

# EFFECTIVE EQUIDISTRIBUTION FOR CLOSED ORBITS OF SEMISIMPLE GROUPS ON HOMOGENEOUS SPACES.

M. EINSIEDLER, G. MARGULIS, A. VENKATESH

ABSTRACT. We prove effective equidistribution, with polynomial rate, for large closed orbits of semisimple groups on homogeneous spaces, under certain technical restrictions (notably, the acting group should have finite centralizer in the ambient group). The proofs make extensive use of spectral gaps, and also of a closing lemma for such actions.

## CONTENTS

1. Introduction.	2
2. A proof of measure classification for semisimple groups.	12
3. Notation and first facts.	16
4. Outline of the proof of Theorem 1.3.	22
5. Basic properties of Sobolev norms.	29
6. Basic properties pertaining to spectral gap.	31
7. Effective generation of Lie algebras.	39
8. Almost invariance of measures.	42
9. Effective ergodic theorem.	44
10. Nearby generic points effectively give additional invariance.	47
11. The measure of points near periodic orbits is small.	51
12. Some lemmas connected to lattice point counting.	53
13. Effective closing Lemma.	55
14. A corollary of the closing Lemma.	60
15. An almost invariant measure on a closed orbit is close to the invariant measure.	65
16. Proof of Theorem 1.3	66
17. An arithmetic application: distribution of integral points on prehomogeneous hypersurfaces	69
Appendix A. Proof of Lemma 3.4.1	75
Appendix B. Proof of Lemma 3.6.1.	77
Appendix C. Proof of (6.1): uniform spectral gap implies $L^p$ coefficients.	79
References	81

## 1. INTRODUCTION.

**1.1. General introduction.** Let  $G$  be a real Lie group, and let  $\Gamma$  be a lattice in  $G$ . Let  $U$  be a connected unipotent subgroup of  $G$ . A theorem of Ratner [56], which proved a conjecture of S. G. Dani, classifies the  $U$ -invariant ergodic probability measures  $\mu$  on  $X = \Gamma \backslash G$ : namely, they are  $S$ -invariant probability measures supported on closed  $S$ -orbits, where  $S$  is a closed subgroup containing  $U$ .

This theorem is a fundamental result with numerous applications. In particular, in combination with results on nondivergence of unipotent orbits, and with the linearization technique, Mozes and Shah [45] proved that the nonzero weak\*-limits of  $U$ -invariant ergodic probability measures are again ergodic, and therefore are Haar measures on closed orbits.

Let  $H \subset G$  be a semisimple subgroup generated by unipotent elements. A special case of the results of [45] asserts that:

(1.1)

The only nonzero weak\*-limits of normalized Haar measures on closed  $H$ -orbits are Haar measures on closed orbits of a closed subgroup  $S \supset H$ .

In the present paper we present a *polynomially effective* version of (1.1) under the assumptions that  $G$  is semisimple,  $\Gamma$  is an arithmetic lattice and  $\text{Lie}(H)$  is centralizer-free inside  $\text{Lie}(G)$ . In explicit terms, we show that the Haar measure on a closed  $H$ -orbit is close to the Haar measure on a closed  $S$ -orbit for a closed subgroup  $S \supset H$ . Moreover, the notions of “close” improves as a power of the volume of the relevant orbits.

We observe that, often, “direct” effectivizations of ergodic proofs only give logarithmically poor (or worse) rates of convergence. The content of our result is, therefore, in the polynomial rate.

A key input is used from the theory of automorphic forms: namely, the  $H$ -action on  $L^2(\mu)$  has a *spectral gap* (if  $\Gamma$  is as in (1) below, and  $\mu$  is the  $H$ -invariant probability measure on a closed  $H$ -orbit). This implies an effective version of the ergodic theorem for  $H$ -actions.

We say that a unitary representation of a (connected Lie) group  $G$  possesses a spectral gap if there exists a compactly supported probability measure on  $G$ , so that convolution with this measure has operator norm strictly less than 1. For a detailed discussion of this concept for semisimple real Lie groups, see §6.

**1.2. Some technical assumptions.** We shall assume that:

- (1) There is a semisimple  $\mathbb{Q}$ -group  $\mathbf{G}$  so that  $G = \mathbf{G}(\mathbb{R})$  and  $\Gamma$  is a congruence subgroup of  $\mathbf{G}(\mathbb{Q})$ ;
- (2)  $H^+ = H$ , i.e.  $H$  has no compact factors and is connected;
- (3) The centralizer of  $\mathfrak{h} = \text{Lie}(H)$  in  $\mathfrak{g} = \text{Lie}(G)$  is trivial.

The third assumption means that closed  $H$ -orbits are “rigid,” i.e., do not come in continuous families. Moreover, it implies (Lemma 3.4.1) that there

are only finitely many intermediate subgroups  $H \subseteq S \subseteq G$ , each of which is semisimple (i.e. the Lie algebra  $\mathfrak{s}$  is a semisimple real Lie algebra).

See §1.8 for a discussion concerning lifting these assumptions. In particular, “congruence” may be weakened to “arithmetic”, at the cost of allowing the exponents in the theorem to depend on the lattice. The assumption of finite centralizer is primarily to “rigidify” the situation; while it does not seem essential to the method, it is clear that many new complications arise when it is removed.

Examples:

- (1)  $G = \mathrm{SL}_n(\mathbb{R})$ ,  $H = \mathrm{SL}_2(\mathbb{R})$  embedded via the irreducible representation. If  $n \geq 4$ , there exists intermediate subgroups  $H \subsetneq S \subsetneq G$ : for instance, if  $n$  is even, the  $H$ -action on  $\mathbb{R}^n$  preserves a symplectic form, and one may take for  $S$  its stabilizer.
- (2)  $G = \mathrm{SL}_n(\mathbb{R})$ ,  $H = \mathrm{SO}(q)^{(0)}$  for some indefinite quadratic form  $q$  in  $n$  variables. In this case, there do not exist intermediate subgroups besides those with connected component  $H$ . (See §17 for an arithmetic application of the theorem in this setting).
- (3)  $G = \mathrm{SL}_2(\mathbb{R})^n$ ,  $H = \mathrm{SL}_2(\mathbb{R})$  embedded diagonally. Here there are many intermediate subgroups as long as  $n \geq 3$ .
- (4)  $G = \mathrm{SL}_7(\mathbb{R})$ ,  $H = G_2(\mathbb{R})$  the group of  $\mathbb{R}$ -points of the  $\mathbb{R}$ -split form of  $G_2$  embedded into  $G$ . Here  $\mathrm{SO}_7$  is an intermediate subgroup. (See §17 for an arithmetic application of the theorem in this setting).

**1.3. Theorem.** *Let  $\Gamma, H \subset G$  be as above satisfying the assumptions in §1.2. Let  $\mu$  be the  $H$ -invariant probability measure on a closed  $H$ -orbit  $x_0H$  inside  $X = \Gamma \backslash G$ .*

*There exists  $\delta, d > 0$  depending only on  $G, H$  and  $V_0 > 0$  depending only on  $\Gamma, G, H$  with the following properties. For any  $V \geq V_0$  there exists an intermediate subgroup  $H \subseteq S \subseteq G$  for which  $x_0S$  is a closed  $S$ -orbit with volume  $\leq V$  and such that  $\mu$  is  $V^{-\delta}$ -close to  $\mu_{x_0S}$ , i.e. for any  $f \in C_c^\infty(X)$  we have*

$$(1.2) \quad \left| \int f d\mu - \int f d\mu_{x_0S} \right| < V^{-\delta} \mathcal{S}_d(f),$$

where  $\mathcal{S}_d(f)$  denotes an  $L^2$ -Sobolev norm of degree  $d$  (see (3.8)).

Let us make certain comments on this result.

- The Sobolev norms  $\mathcal{S}_d(f)$  measure the  $L^2$ -norm of the derivatives of  $f$  up to the  $d$ -th order w.r.t. a smooth measure on  $X$ . If  $X$  is compact, this defines them up to bounded multiples.

An equivalent way of formulating (1.2), without the use of Sobolev norms, would be in terms of the measure of small balls. For simplicity, consider the case when  $X$  is compact and fix a Riemannian metric on  $X$ . There exists  $\delta_1, V_1 > 0$  so that, for any  $V \geq V_1$  and for any Riemannian ball  $B(r)$  of radius  $r \leq 1$ , we have

$$|\mu(B(r)) - \mu_{x_0S}(B(r))| \leq V^{-\delta_1}$$

Such an assertion is equivalent to (1.2), up to modification of constants.

- In interpreting the theorem – which at first sight is rather confusing owing to the parameter  $V$  – it may be helpful to understand how it implies the *non-effective* variant, viz. theorem of Mozes and Shah (1.1).

We take for granted the following fact, for any intermediate subgroup  $H \subset S \subset G$ :

(1.3)      There are finitely many closed  $S$ -orbits with volume  $\leq V$ .

An effective version of (1.3), with polynomial bounds, is given in §11, but the qualitative statement is simple to establish.

Let  $X = \Gamma \backslash G$  be as in Theorem 1.3. Suppose we are given a sequence  $x_i H$  of closed  $H$ -orbits so that the sequence  $\mu_{x_i H}$  converges to a weak\*-limit  $\nu_\infty$ .

Fix, for a moment, an intermediate subgroup  $S \supset H$ . Let us agree to set  $\text{vol}(x_i S) = \infty$  if  $x_i S$  is not closed. Passing to a subsequence, we may assume that either there exists some  $M > 0$  such that the volume of  $x_i S$  is less than  $M$  for all  $i$ , or the volume of  $x_i S$  approaches infinity as  $i \rightarrow \infty$ .

Since there are finitely many intermediate subgroups, we may assume, passing to a further subsequence, that one or the other option holds for all such  $S$ . That is to say: there is a subset  $\mathcal{S}$  of the set of intermediate subgroups  $H \subset S \subset G$  so that  $x_i S$  is closed and of volume  $\leq M$  when  $S \in \mathcal{S}$ ; and of volume approaching  $\infty$ , when  $S \notin \mathcal{S}$ .

Passing to a further subsequence, and applying (1.3), we may suppose that, for any  $S \in \mathcal{S}$ , there exists  $y_S$  so that  $x_i S = y_S S$  for all  $i$ . Apply the theorem with fixed  $V > M$ . It follows that  $\nu_\infty$  is  $V^{-\delta}$ -close to a measure of the form  $\mu_{y_S S}$  for some  $S \in \mathcal{S}$ . Since  $V$  was arbitrary, our conclusion follows.

- Our proof gives the same result for any  $H$ -invariant measure  $\mu$  so that the  $H$ -action on  $L^2(\mu)$  possesses a spectral gap. In other terms, the true content of the theorem is the “effective classification of *strongly* ergodic  $H$ -invariant measures.” The only naturally arising context, however, seems to be the case described in the theorem.
- *In principle*, the exponent  $\delta$  is computable; however, it would be rather painful to compute and extremely small in general.

In the case when  $H$  is maximal, it is likely not too difficult to compute and not too small. However, our goal has been only to achieve polynomial dependence. This is of qualitative significance, for “abstract” effectivizations of results in ergodic theory yield, in general, far poorer bounds.

- The theorem also has a purely topological version that we now enunciate.

**1.4. Theorem.** *Notations as in Theorem 1.3, let  $F \subset X$  be compact, and fix a Riemannian metric on  $X$ . There exists  $\rho > 0$ , depending on  $G$ ,  $H$ , and  $V_1$  depending on  $\Gamma$ ,  $G$ ,  $H$ ,  $F$ , and the choice of Riemannian metric, with the following property:*

*For any  $V \geq V_1$ , there exists  $S$  so that  $x_0S$  is a closed orbit of volume  $\leq V$ , and any point in  $x_0S \cap F$  is at distance  $\leq V^{-\rho}$  from  $x_0H$ , w.r.t. the Riemannian metric induced on  $x_0S$  from  $X$ .*

By the *Riemannian metric induced on  $x_0S$* , we mean: endow  $x_0S$  with the Riemannian structure induced as a submanifold of  $X$ , and compute the corresponding metric.

Let us note in particular the import of Theorem 1.4 in the case when  $H$  is a *maximal* subgroup in  $G$ . Then there exists  $\ell \geq 1$  so that a closed  $H$ -orbit of volume  $\geq \varepsilon^{-\ell}$  has to intersect any  $\varepsilon$ -ball in  $X$ . In particular, we also prove that the number of closed  $H$ -orbits that fail to intersect some  $\varepsilon$ -ball is bounded by a polynomial in  $\varepsilon^{-1}$ .

**1.5. Outline of the proof.** First of all, let us consider the *non-effective* version of Theorem 1.3, i.e. (1.1). This special case of measure rigidity (when the acting group is semisimple) admits a simple proof. We present it in §2. One may then see the proof of Theorem 1.3 as an effective form of that proof.<sup>1</sup> We suggest the reader begin by reading this non-effective proof.

In §4, we present a detailed outline of the proof of the main Theorem, comparing each portion of the proof to the corresponding non-effective statement in §2. The reader may wish to read §4 after §2.

For now we present a very brief high-level outline of the proof strategy: We show for increasingly large intermediate subgroups  $H \subseteq S \subseteq G$ , *either*  $\mu$  is supported on an  $S$ -orbit of small volume, *or*  $\mu$  is almost invariant by a strictly larger subgroup  $S' \supset S$ .

Indeed, take a one-parameter unipotent subgroup  $U \subset H$  and say a point  $x \in X$  is generic if the measure along long segments of  $x.U$  approximates well the measure  $\mu$ . The proof of the dichotomy uses the following principles:

- (1) (Ergodic theorem). Most points  $x \in X$  are generic.
- (2) (Nearby generic points give additional invariance). If  $x, x' \in X$  are two generic points which are very close, i.e.  $x' = xg$  for some  $g$  near the identity, and  $g \notin S$ , then  $\mu$  is almost invariant by a larger subgroup  $S' \subset S$ .
- (3) (Dichotomy). If we *cannot* find  $x, x'$  as above, then necessarily  $\mu$  was supported on a closed  $S$ -orbit of small volume.

Step (2) is easily effectivizable, relying only on polynomial behavior of unipotent trajectories (see §1.7.1).

---

<sup>1</sup>In this case, the process of *effectivization* causes proofs to increase in length approximately twenty-fold!

“Good” effective forms of both (1) and (3), however, make use of spectral gaps. Specifically, we use the fact that, for any intermediate subgroup  $S \supset H$ , the action of  $S$  on  $L_0^2(xS)$ , where  $xS$  varies through all closed  $S$ -orbits, has a uniform spectral gap. (Here  $L_0^2$  denotes the orthogonal complement of locally constant functions.) The crucial ingredient in providing the spectral gaps we need is the work of Clozel on “property  $\tau$ .” In effect, this guarantees<sup>2</sup> a uniform spectral gap for  $Q$  acting on  $L^2(\Lambda \backslash Q)$ , when  $Q$  is a semisimple group and  $\Lambda$  varies through *congruence* lattices within  $Q$ . (See also §1.8.2.)

We also make use of several techniques that have become, to some extent, standard in the theory of homogeneous dynamics. Most notably we use (as noted) polynomial divergence for unipotent actions and the related linearization argument, which we will discuss in §1.7.1. These ideas will enter both in steps (2) and (3).

Finally, the process of “effectivizing” results concerning invariant measures, inevitably, leads to questions about efficient generation of Lie groups; in handling these (technical) issues, we shall make use of some simple arguments in algebraic geometry and Diophantine geometry, as well as of the Łojasiewicz inequality.

**1.6. Organization of the paper.** The paper is organized as follows:

We start in §2 with the non effective version of Theorem 1.3 and the simplified proof of Ratner’s theorem on measure rigidity in the case of semisimple acting groups.

In §3, we set up notation. Some of this may be of independent interest; our use of Sobolev norms and relative traces, inspired by a paper of Bernstein and Reznikov, seems a natural way of handling certain analytic questions on homogeneous spaces.

In §4, we give a detailed sketch of the proof of the Theorem. We give precise statements of the main constituent results that go into the Theorem, and references to their proofs in the text. We do not give proofs, but we indicate the main idea in each case.

In §5, 6, 7, 8 respectively, we establish or recall some basic properties pertaining to, respectively, Sobolev norms, the spectral gap, generation of Lie algebras, and almost invariance of measures. The reader might wish to skip these sections and refer to them as necessary while reading the rest of the text.

In the sections §9 – §14, we prove the central statements required for the Theorem.

§16 gives the proof of Theorem 1.3 and its corollaries.

§17 gives a simple application of the Theorem to a number-theory problem.

---

<sup>2</sup>The terminology “property  $\tau$ ” is perhaps confusing. In general, it asserts that group  $Q$  has a uniform spectral gap in its action on some natural family of unitary representation. Therefore, unlike property  $T$ , which is an intrinsic property of a group  $G$ , property  $\tau$  only makes sense by reference to an implicit choice of such a family.

The Appendices recall and prove certain basic results that we need.

### 1.7. Connection to existing work.

1.7.1. *Dynamics of unipotent flows.* Given a one parameter subgroup  $g_t \in G$ , and two “nearby” points  $x_1, x_2 \in X$ , satisfying  $x_2 = x_1 \exp(v)$ , for  $v \in \text{Lie}(G)$ , we have

$$x_2 g_t = x_1 g_t \exp(\text{Ad}(g_{-t})v).$$

The behavior of  $t \mapsto \text{Ad}(g_{-t})v$  controls the relative behavior of the orbits of  $x_1$  and  $x_2$  under  $g_t$ . In general,  $t \mapsto \text{Ad}(g_{-t})v$  may grow exponentially. However, if  $g_t$  happens to be a *unipotent* subgroup, then  $t \mapsto \text{Ad}(g_{-t})v$  is a polynomial. This feature is the *polynomial behavior* or *polynomial divergence* of unipotent flows.

The polynomial divergence of orbits of unipotent flows was one of the main motivations for M.S. Raghunathan when he formulated, in the mid-seventies, his conjecture about “orbit closure rigidity.” He hoped that unipotent flows are likely to have “manageable behavior” because of the slow divergence of orbits of unipotent one-parameter subgroups (in contrast to the exponential divergence of orbits of diagonalizable subgroups).

The polynomial behavior of unipotent orbits is very important in the present paper, and we have used many techniques developed over the past four decades. We make particular note of the following three ideas we use; this is not intended as either a survey or a history of the field.

1. Nondivergence of unipotent flows. One of the first times the polynomial behavior was used was in the proof of non-divergence of unipotent flows by G.M. [36], where it was indeed one of the basic ingredients. This phenomenon of non-divergence was quantified and further refined by Dani, Kleinbock, and G.M. [10, 13, 27]. We actually make use of [27] to control how much mass of a closed  $H$ -orbit can be close to infinity, see Lemma 3.6.1.

2. “Nearby generic points give additional invariance.” A topological incarnation of this principle was utilized in the work of G.M. and G.M.-Dani [11, 12, 14, 38, 39, 40], and a measure-theoretic incarnation was utilized in the work of M. Ratner in [53, 54, 55, 56]; the former uses Chevalley’s theorem in order to control relative behavior of orbits of unipotent flows, and the arguments of the latter papers utilize the “R-principle,” generalizing the “H-principle” used earlier by Ratner [49, 50, 51, 52] and by Witte [64, 65].

To be a little more precise, the arguments of [11, 12, 14, 38, 39, 40] show that, if  $U$  is a unipotent subgroup of  $G$ , and  $Y$  is a minimal closed  $U$ -invariant subset of  $\Gamma \backslash G$ , then either:

- (1)  $Y$  is an orbit of  $U$ , or
- (2)  $Y$  is invariant under a bigger connected subgroup contained in the normalizer of  $U$  (additional invariance).

Arguments contained in [53, 54, 55, 56] show that, if  $\mu$  is an ergodic  $U$ -invariant measure, then either:

- (1)  $\mu$  is the Haar measure on a closed  $U$ -orbit, or
- (2)  $\mu$  is invariant under a bigger connected subgroup contained in the normalizer of  $U$  (additional invariance).

The proof of these results may be understood as topological and measure-theoretic versions of the principle enunciated above; the notion of *minimal set* is the topological analog of the notion of *ergodic measure*. Indeed, the proof of the topological statement is based on studying the orbits of a sequence of points that converge to a minimal closed invariant set; the proof of the measure-theoretic statement is based on studying the orbits of nearby points that are generic in the sense of Birkhoff's ergodic theorem.

Let us draw attention to one important difference between the topological and the measure-theoretic setting: an arbitrary  $U$ -invariant measure may be decomposed into ergodic components; the analog of this statement for a  $U$ -invariant closed set is not true.

Our use of an effective ergodic theorem, together with polynomial divergence, is motivated by this principle; many of our arguments are closely related to (ineffective) arguments from the papers mentioned above. These ideas enter, in particular, into steps (1) and (2) of the outline of the proof presented in §1.5.

3. The “linearization” technique. By this, we shall mean a collection of methods which allow one to control the time an orbit of a unipotent subgroup  $U$  spends near closed orbits of certain subgroups  $S \subset G$  containing  $U$ . One method to do this is by realizing the action of  $U$  on  $G/S$  as the linear action of  $U$  on a certain subset of a linear space (which also makes use of Chevalley's theorem). The origins of this method can be traced back to the work of S.G.Dani and Smillie [15] in the context of Fuchsian groups. General results and techniques are developed in [11, 12, 14, 38, 39, 40, 42, 53, 54, 55, 56].

We make use of a version of this technique to control the time spent near a closed orbit of a semisimple subgroup, see Proposition 11.1; this enters into step (3) of the outline of §1.5.

1.7.2. *Effective equidistribution.* There are two general cases where it is known one may establish effective results about the distribution of  $H$ -orbits on  $\Gamma \backslash G$ :

- (1) When  $H$  is a *horospherical* subgroup, i.e. the Lie algebra of  $\mathfrak{h}$  consists of all contracted directions for the adjoint action of a semisimple element  $s \in G$ . In that case, one may use the mixing properties of the  $s$ -action on  $\Gamma \backslash G$ .
- (2) When  $H$  is “large” inside  $G$ , one may sometimes analyze effectively the distribution of closed  $H$ -orbits inside  $G$  via representation theory or automorphic forms. “Large” usually means, *at the very least*, that  $H$  should act with an open orbit on the flag variety  $G/P_0$ , where  $P_0$  is a minimal parabolic of  $G$ .

In the case (1), the approach using mixing properties of the  $s$ -action on  $\Gamma \backslash G$  can be traced back to the thesis of G.M. [41], where it was used in

the context of Anosov flows. (This thesis was written in 1970, but published more than thirty years later). Another approach to effective equidistribution for long horocycles on  $\mathrm{SL}_2(\mathbb{R})$  is implicit in the work of Ratner [52, 57], where the effective ergodicity of the horocycle flow is used. Both of these ideas would suffice to prove effective equidistribution of orbits of horospherical subgroups in any rank, as would the work of Burger [4]. More detailed analysis of the quantitative equidistribution of the horocycle flow for  $\mathrm{SL}_2(\mathbb{R})$  may be found in [23, 60]; quantitative mixing rates are discussed in [26], and analysis of equidistribution of closed horospheres and its relevance to the theory of automorphic forms may be found in [61].

A typical instance of the second type of result is the equidistribution of Hecke points, [22] or [7]; a “twisted” version is the work of Li, Jiang and Zhang, [33]. These both correspond to the special case of the main result when  $H$  is diagonally embedded in  $G = H \times H$ ; in the “Hecke point” cases, one restricts the possibilities for the  $\mathbb{Q}$ -form underlying the closed  $H$ -orbit, whereas the latter case restricts to  $H = \mathrm{PGL}_2$  over a totally real number field.

An illuminating context where the two cases overlap is the question of *equidistribution of translates* for a closed orbit of a symmetric subgroup: see [16] for an effective treatment by the second method, and [1] for an effective treatment in a more general setting, following the strategy of [19], by a method very closely related to the first. (The “wavefront lemma” assures that, in the limit, such translates acquire local invariance by a horospherical subgroup).

In both cases above, the analysis usually makes use of spectral gaps; for instance, the first method makes use of quantitative mixing, which is substantially equivalent to a spectral gap. Spectral gaps are also used in our argument, but in an essentially different way; most importantly, we do not use a spectral gap for the ambient group  $G$  acting on the ambient space, but rather for the acting group  $H \subset G$  acting on the invariant measure. The idea that this could be used to give a simple proof – in certain cases – of the fact that “limits of ergodic measures remain ergodic” was used in a paper by J. Ellenberg and A.V. [20], based on prior discussions with M. E. and Elon Lindenstrauss.

In any case, the methods in both cases are fundamentally limited even if  $H$  is a maximal subgroup of  $G$ . Moreover, these methods do not detect closed orbits of intermediate subgroups  $H \subset S \subset G$ , and so there appears to be little hope of generalizing them significantly. Indeed, the result of this paper appears to be the first that produces quantitative results when  $G$  is semisimple and  $H$  is not one of the subgroups mentioned in (1) or (2) above, and, in particular, in the case when  $H$  is “far from maximal.”

**1.7.3. A history of this paper.** A topological version of Theorem 1.3 was proven by G.M., in an unpublished manuscript, in the case when  $H =$

$\mathrm{SO}(2, 1)$  and  $G = \mathrm{SL}_3(\mathbb{R})$  (and implicitly in the case when  $H$  is a maximal semisimple subgroup of  $G$ ).

The present paper uses a different technique than the one used in that proof. In the case of  $H = \mathrm{SO}(2, 1)$ ,  $G = \mathrm{SL}_3(\mathbb{R})$ , that proof uses mixing properties of the action of a torus within  $H$ , whereas the proof presented in this paper uses ergodic properties of a one-parameter unipotent flow within  $H$ . A discussion of the relationship of these techniques is presented in §2.

The idea of using effective ergodic theorems for the unipotent flow was noted independently by M.E. and A.V., and by G.M., motivated by various methods in the theory of unipotent flows.

We note that the general case is very considerably more involved than the case of  $\mathrm{SO}(2, 1) \subset \mathrm{SL}_3(\mathbb{R})$ , owing to the possibility of intermediate subgroups. In fact, the technically most complicated arguments, e.g. the effective closing lemma, are necessary only if  $H$  is not maximal.

## 1.8. Remarks on generalizations.

1.8.1. *The congruence subgroup assumption.* One may replace the assumption that  $\Gamma$  is congruence by the assumption that  $\Gamma$  is arithmetic. We indicate here how this is done.

Suppose that  $\Gamma$  is an arithmetic, but not necessarily congruence, subgroup of  $\mathbf{G}(\mathbb{Q})$  and  $xH$  is a closed  $H$ -orbit on  $X = \Gamma \backslash G$ . Let  $\Lambda$  be a congruence subgroup of  $\mathbf{G}(\mathbb{Q})$ . Replacing  $\Gamma$  with  $\Gamma \cap \Lambda$  and  $xH$  by the orbit  $x'H$  of a preimage  $x'$  of  $x$  in  $\Gamma \cap \Lambda \backslash G$ , we may assume that  $\Gamma \subset \Lambda$ . Let  $\bar{X} = \Lambda \backslash G$ , so we have a natural projection  $\pi : X \rightarrow \bar{X}$ .

Let  $\mu$  resp.  $\bar{\mu}$  be the  $H$ -invariant measures on  $xH$  resp. their projections to  $\bar{X}$ . Our main theorem gives an effective result about the distribution of  $\bar{\mu}$ . Suppose, for simplicity, we are in the situation where  $\bar{\mu}$  is close to the  $G$ -invariant probability measure  $\mu_{\bar{X}}$  on  $\bar{X}$ . Here the error of approximation is quantified by Theorem 1.3.

Let  $f \in C_c^\infty(X)$  be so that  $\mu_X(f) = 0$ . Let  $u(t)$  be a one-parameter subgroup in  $H$ , and let  $f_T := \frac{1}{T} \int_{t=0}^T u(t)f$ .

For a function  $f$  on  $X$ , we define  $\pi_* f$  on  $\bar{X}$  via  $\pi_* f(\bar{x}) = \sum_{\pi(x)=\bar{x}} f(x)$ .

Then

$$|\mu(f)|^2 = |\mu(f_T)|^2 \leq \mu(|f_T|^2) \leq \bar{\mu}(\pi_* |f_T|^2) \sim \mu_{\bar{X}}(\pi_* |f_T|^2) = [\Lambda : \Gamma] \mu_X(|f_T|^2)$$

where the rate of approximation depends on the quality of approximation of  $\mu_{\bar{X}}$  by  $\bar{\mu}$  and on the Sobolev norm of  $\pi_* |f_T|^2$  and so also on  $T$ . The right-hand quantity is bounded by quantitative mixing of the  $G$ -action on  $X$ . Optimizing for  $T$  shows that  $\mu_X(f) = 0$  implies that  $|\mu(f)|$  is “small”, which is the same as  $\mu \sim \mu_X$ .

The above argument is entirely quantitative. The quality of the bound depends on the spectral gap for  $G$  acting on  $L^2(X)$ .

Observe an important difference with Theorem 1.3: in the case when  $\Gamma$  is assumed only to be arithmetic, the quality of the bound depends on the lattice  $\Gamma$ , and not only on  $G, H$ . Presumably, such a dependence is not avoidable, as one may see by considering the example of the horocycle flow on arithmetic quotients of  $\mathrm{SL}_2(\mathbb{R})$ .

**1.8.2. *S*-arithmetic generalization. Translates of a fixed orbit.** One may envisage a version of the theorem that concerns closed orbits of a semisimple *S*-arithmetic group  $H$  acting on an *S*-arithmetic quotient  $\Gamma \backslash G$ .

This extension has more than simple formal significance. We indicate three applications which do not on their face involve *S*-arithmetic groups.

- (1) It is possible to give an independent proof of property  $(\tau)$  (first established by Clozel [6]) using the *S*-arithmetic extension of our theorem.<sup>3</sup>
- (2) Notations as in our main theorem, let  $x_0H$  be a closed  $H$ -orbit. One may ask about the distribution of *varying translates* of  $H$ , i.e.  $x_0Hg$  when  $g \rightarrow \infty$ . If  $H$  is a symmetric subgroup, effective distribution results may be given using the wavefront lemma, cf. [19]. We anticipate that the *S*-arithmetic extension of our result will allow us to treat this question in the more general setting of §1.2.
- (3) In an arithmetic direction, [20] uses the case when  $H, G$  are  $p$ -adic orthogonal groups to prove local-global principles for representations of quadratic forms. To establish an effective version of that result would imply new bounds for the Fourier coefficients of Siegel modular forms. Results in this direction would require removal of the centralizer assumption from §1.2.

As is usual in such matters, the existence of small additive subgroups of  $\mathbb{Q}_p$  will cause further complications.

**1.8.3. *The centralizer assumption.*** There exist many natural settings where the centralizer assumption of §1.2, assumed in Theorem 1.3, is too restrictive. Indeed, it does not seem to be truly essential to our method; the key part of our method is the existence and exploitation of spectral gaps.

However, many technical complications seem to arise when it is removed. We hope to discuss this elsewhere.

**1.9. Acknowledgements.** This work was initiated at the Institute of Advanced Study during the academic year 2005-2006; in fact, all authors were visiting there for various parts of that year. We would like to thank the IAS for providing excellent working conditions.<sup>4</sup>

<sup>3</sup>Note that, in the present paper, we use property  $(\tau)$  as an input. However, it is possible to avoid it by more elaborate arguments, and indeed to *derive* it in general by our methods, at least for groups of absolute rank  $\geq 2$ . The groups of absolute rank 1 require an alternative treatment; this is also true of Clozel's approach.

<sup>4</sup>The first and the third named authors would like to make particular note of the excellence of the cookies.

This research has been supported by the Clay Mathematics Institute (M.E. by a Clay Research Scholarship, A.V. by a fellowship); and by the NSF (DMS grants 0622397 (M.E.), 02045606 (A.V.), (0244406) (D.M.), and an FRG collaborative grant 0554373 (M.E. & A.V.))

We would like to express our appreciation of conversations with E. Lapid (concerning the proof of (6.1)) and H. Oh and Y. Shalom (concerning §17 and §2.9).

We would also like to express our gratitude to Elon Lindenstrauss, who both contributed greatly to early discussions about this paper, and encouraged its completion. It is a great pleasure to thank him for his generosity with ideas and time.

## 2. A PROOF OF MEASURE CLASSIFICATION FOR SEMISIMPLE GROUPS.

**2.1. Introduction.** In the present section, we give a short proof of the non-effective version of Theorem 1.3, the result (1.1) of Mozes and Shah. Notation as in the theorem, this proof has three (independent) constituents:

- (1) Proposition 2.2 classifies  $H$ -ergodic measures on  $\Gamma \backslash G$ ;
- (2) Proposition 2.10 shows that, under a certain spectral gap assumption, any limit of  $H$ -ergodic probability measures on  $\Gamma \backslash G$  is itself ergodic.
- (3) It is established in Proposition 6.7.1 that this spectral gap assumption is valid in the setting of (1.1).

Therefore, taken in combination, these results establish (1.1).<sup>5</sup>

While the proofs of Proposition 2.2 and Proposition 2.10 are elementary, Proposition 6.7.1 is a deep result from the theory of automorphic forms. The key virtue of this proof, however, is amenable to effectivization. Indeed, the contents of this section may be viewed as the non-effective counterpart of the proof of Theorem 1.3. §4 gives an outline of the effective proof of Theorem 1.3, together with comparisons to the non-effective proof presented here.

**2.2. Proposition.** *Let  $G$  be a semisimple real Lie group, let  $\Gamma$  be a discrete subgroup in  $G$ , and let  $H$  be a connected semisimple subgroup of  $G$  without compact factors. Let  $\mu$  be an  $H$ -invariant and ergodic probability measure on  $X = \Gamma \backslash G$ . Then  $\mu$  is the  $S$ -invariant probability measure on a closed  $S$ -orbit of an intermediate subgroup  $H \subseteq S \subseteq G$ .*

This is due to M. Ratner [56]. In the setting considered here, where  $H$  is semisimple, the proof simplifies significantly. This has been noted already in [17] due to M.E. We shall present a simplified proof along these lines,

---

<sup>5</sup>To be precise, in the case where  $\Gamma \backslash G$  is noncompact, there may be “escape of mass,” i.e., it is not a priori obvious that the limit of such a sequence is a probability measure. However, in the case considered in Theorem 1.3 –  $\mathfrak{h}$  has trivial centralizer – this does not occur: Lemma 3.6.1.

differing somewhat from [17]. See also [55, p244] where similar arguments are presented.

Just as the outline presented in §1.5, the proof of Proposition 2.2 has three distinct steps:

- (1) *Ergodic theorem*: §2.4.
- (2) *Nearby generic points give additional invariance*: §2.5.
- (3) *Dichotomy*: Either one can find nearby generic points, or  $\mu$  is supported on a closed orbit. §2.6.

**2.3. Setup for the proof.** We now indicate the proof of Proposition 2.2. Let  $G, H, \Gamma, X = \Gamma \backslash G$ , and  $\mu$  be as in Proposition 2.2.

Let  $S = \{g \in G : g \text{ preserves } \mu\}$ . The subgroup  $S$  is closed and contains  $H$ .  $H$  being semisimple, the restriction to  $H$  of the adjoint representation of  $G$  is completely reducible. Hence there exists an  $\text{Ad}(H)$ -invariant complement  $\mathfrak{r}$  to the Lie algebra  $\mathfrak{s}$  of  $S$ , within  $\mathfrak{g}$ . It is the precisely the existence of this  $\text{Ad}(H)$ -invariant complement which simplifies the proof of the measure classification result.

**2.4. Ergodic theorem.** Let  $U = \{u(t) : t \in \mathbb{R}\}$  be a one-parameter unipotent subgroup of  $H$  which projects nontrivially on each simple factor of  $H$ . Then, according to a theorem of Moore (Mautner phenomenon) the measure  $\mu$  is  $U$ -ergodic.

According to Birkhoff's individual ergodic theorem, for  $\mu$ -almost  $x \in X$  and any continuous compactly supported function  $f$  on  $X$ ,

$$(2.1) \quad \frac{1}{T} \int_0^T f(xu_{-t}) dt \rightarrow \int f d\mu.$$

A point  $x \in X$  satisfying (2.1) for all  $f \in C_c(X)$  is called *generic*.

We say that  $E \subset X$  is a set of uniform convergence if for any compactly supported function  $f$  the above convergence is uniform with respect to  $x \in E$ . In view of Egoroff's theorem, there exist sets of uniform convergence of measure arbitrarily close to 1. Fix such a set  $E$  of uniform convergence of  $\mu$ -measure  $> 9/10$ .

**2.5. Nearby generic points give additional invariance.** First let us make the following remark, which is a quantification of "polynomial behavior" discussed in §1.7.1:

- (i) For any element  $g$  of  $G$ , the matrix coefficients of  $\text{Ad}(u_t g u_{-t})$  are polynomials in  $t$  of degree not greater than  $\dim(G)$ .

Let  $B$  be an open bounded subset of  $G$  containing the identity. For  $g \in B$ , we set

$$\begin{aligned} T_g &= \sup \{T : u_t g u_{-t} \in B \text{ whenever } t \in (0, T)\}, \\ g^* &= u_{T_g} g u_{-T_g}. \end{aligned}$$

It follows from (i) that  $T_g$  is finite if  $g$  does not belong to the centralizer  $C_G(U)$  of  $U$  in  $G$ . Moreover, in that case  $g^*$  is well defined and belongs to the boundary of  $B$ .

For  $g \in B \setminus C(U)$ , we define a map  $q_g : [0, 1] \rightarrow G$  by

$$q_g(s/T_g) = u_s g u_{-s}, s \in [0, T_g].$$

Since  $B$  is bounded, the family  $\{q_g\}$  is uniformly bounded. Therefore, the adjoint actions of these elements form an equicontinuous family of polynomial maps on the Lie algebra. This implies the following statement, for a metric  $d$  on  $X$  that is obtained from a left-invariant metric on  $G$ :

- (ii) For every  $\epsilon > 0$  one can find  $\delta > 0$  such that if  $x \in X$ ,  $g \in B \setminus C_G(U)$ , and  $t \in [(1 - \delta)T_g, T_g]$ , then the distance

$$(2.2) \quad d(xgu_{-t}, xu_{-t}g^*) \leq d(u_t g u_{-t}, g^*) < \epsilon.$$

**2.5.1. Lemma.** (a) Let  $g$  be an element of  $C_G(U)$ . Suppose that we can find  $x, y \in E$  such that  $y = xg$ . Then the measure  $\mu$  is invariant under  $g$ .

(b) Suppose that we can find a sequence  $g_n \in G$  and a sequence  $x_n \in E$  such that

- (1)  $g_n \rightarrow 1$ ;
- (2)  $g_n \notin C_G(U)$  for all  $n$ ,
- (3)  $x_n g_n \in E$  for all  $n$ .

Then  $\mu$  is invariant under any limit point of the  $g_n^*$ .

*Proof.* The first part follows from the genericity of the points.

Consider the second assertion. Let  $g^*$  be a limit point of  $g_n^*$ , which exists by compactness of the closure of  $B$ . Let  $f \in C_c(X)$ . Observe that  $T_{g_n} \rightarrow \infty$ .

Given  $\varepsilon > 0$  and  $\delta > 0$  we may apply the definition of  $E$  (for  $T = T_{g_n}$  and  $T = \delta T_{g_n}$ ), that  $x_n g_n \in E$ , and that  $x_n \in E$  to obtain for all sufficiently large  $n$  that

$$\left| \frac{1}{\delta T_{g_n}} \int_{(1-\delta)T_{g_n}}^{T_{g_n}} f(x_n g_n u_{-t}) dt - \int f d\mu \right| < \varepsilon',$$

$$\left| \frac{1}{\delta T_{g_n}} \int_{(1-\delta)T_{g_n}}^{T_{g_n}} f(x_n u_{-t} g^*) dt - \int f(x g^*) d\mu(x) \right| < \varepsilon'$$

Choosing  $\delta > 0$  as in (ii), depending on  $\varepsilon > 0$  and the uniform continuity of  $f$ , we conclude in the limit that the  $\mu$ -integral of  $x \mapsto f(x)$  and  $x \mapsto f(xg^*)$  coincide.  $\square$

## 2.6. Dichotomy.

**2.6.1. Lemma.** Assume that  $\mu(xS) = 0$  for all  $x \in X$ . Then there exists a sequence  $x_n \in E$  and a sequence  $g_n \in \exp(\mathfrak{t} \setminus \{0\})$  converging to 1 such that  $x_n g_n \in E$ .

*Sketch of the proof.* Let  $O$  be a bounded open neighborhood of the identity in  $S$ , and let  $E_1$  be the set of all points  $x$  in  $E$  such that the relative measure in  $O$  of the set  $\{s \in O : xs \in E\}$  is greater than  $8/10$ . It follows from the Fubini theorem that  $\mu(E_1) > 1/2$ .

For each  $n > 1$  take two points  $y_n, z_n \in E_1$  that satisfy  $d(y_n, z_n) < 1/n$ , but do not satisfy  $y_n = z_n s$  for any  $s \in S$  near the identity. A density computation shows that there exists  $s_n, s'_n \in O$  so that  $y_n s_n \in E, z_n s'_n \in E$ , and  $z_n s'_n = y_n s_n \exp(r_n)$  for some  $r_n \in \mathfrak{r}$ . This is quite standard, see e.g. [17, Lemma 4.4, 4.5] for details.

Take  $x_n = y_n s_n$ , and take  $g_n = \exp(r_n)$ . The resulting sequences have the desired properties.  $\square$

**2.7. Conclusion of the proof.** Suppose that  $\mu(xS) = 0$  for all  $x$ . By using Lemma 2.5.1 and Lemma 2.6.1 we show that  $\mu$  is  $g$ -invariant, for some  $g \in \overline{B} \cap \exp(\mathfrak{r} \setminus \{0\})$ . By choosing  $B$  sufficiently small, we may assume that  $\overline{B} \cap \exp(\mathfrak{r} \setminus \{0\})$  is disjoint from  $S$ . Contradiction.

Thus  $\mu(xS) > 0$  for some  $x \in X$ . The stabilizer  $S_x = \{s \in S : xs = x\}$  of  $x$  is a lattice in  $S$ . Consequently,  $S$  is unimodular and  $xS$  is closed. We are done, for  $\mu$  is ergodic.  $\square$

**2.8. A remark about the two approaches in §1.7.3.** We may phrase the proof of Proposition 2.2 in qualitative forms as follows:

(2.3) The measure along the trajectory  $\{xu(t) : 0 \leq t \leq T\}$   
approximates  $\mu$  for “most”  $x$ .

One then finds two such points  $x_1, x_2$  and studies the relation between their trajectories in order to obtain additional invariance.

(2.3) is a special case of a more general fact, not specific to unipotent trajectories: if  $\nu$  is a measure on  $H$  with “spread-out support,” and we let  $\nu_x$  be the push-forward of  $\nu$  by the map  $h \rightarrow xh$ , then  $\nu_x$  will approximate  $\mu$  for “most”  $x$ ; the previous remark is the special case when  $\nu$  is localized along a long trajectory of  $u(t)$ .

In §1.7.3, we remarked on another method of treating the case of  $H \subset G$  maximal, using mixing properties of semisimple elements. We are now in a position to briefly describe it.

In that approach, one considers instead the measures  $\nu$  that are obtained by translating a fixed, compactly supported density on  $H$  by a large semisimple element. In particular, these measures  $\nu$  resemble large pieces of a coset of a horospherical subgroup of  $H$ . As in the proof with unipotent flows given here, one studies the relation between  $\nu_{x_1}, \nu_{x_2}$  for nearby  $x_1$  and  $x_2$  to conclude additional invariance.

**2.9. Ergodicity of limit measures in presence of a spectral gap.**

**2.10. Proposition.** *Let a  $\sigma$ -compact metric group  $H$  act continuously on a  $\sigma$ -compact metric space  $X$ . Let  $\mu_n$  be a sequence of  $H$ -invariant and ergodic measures so that the  $H$ -action on*

$$\{f \in L^2(\mu_n) : \int f d\mu_n = 0\}$$

*possesses a spectral gap which is uniform in  $n$ . Then any weak\*-limit of the  $\mu_n$  is ergodic under  $H$ .*

This result is due to E. Glasner and B. Weiss, in the case when  $H$  has property (T). More along the lines of the current paper: it is used to give a “cheap” proof of a special case of a Mozes-Shah type result, in a  $p$ -adic setting, in [20].

Note that we do not say whether the weak\* limit is still a probability measure.

For our purposes, it will be important that, in a suitable sense, the proof of Proposition 2.10 is effective. Indeed, fixing a compact generating set  $K \subset H$ , there exists  $\delta > 0$  so that:

$$\sup_{k \in K} \mu_\infty(kZ - Z) \geq \delta \mu_\infty(Z)$$

for any Borel set  $Z \subset X$  with  $\mu(Z) \leq 1/2$ , where  $\mu_\infty$  is any weak\* limit of the sequence  $\mu_n$ .

In other terms, there are no *almost* invariant subsets.

*Proof.* Let  $\nu$  be a compactly supported probability measure on  $H$ , so that for every  $n$ :

$$(2.4) \quad \|\nu \star f - \int f d\mu_n\|_{L^2(\mu_n)} \leq \frac{1}{2} \|f\|_{L^2(\mu_n)}$$

for all  $f \in L^2(\mu_n)$  (or equivalently for all  $f \in C_c(X)$ ). Here  $\nu \star f$  denotes convolution of  $f$  under  $\nu$  w.r.t. the action. The measure  $\nu$  exists by virtue of the assumption on the uniform spectral gap. Passing to the limit, one easily verifies that (2.4) remains valid with  $\mu_n$  replaced by any weak\*-limit thereof. This implies that such a weak\*-limit must be  $H$ -ergodic.  $\square$

### 3. NOTATION AND FIRST FACTS.

Fix, first of all, a semisimple  $\mathbb{Q}$ -group  $\mathbf{G}$  and a Euclidean norm  $\|\cdot\|$  on the Lie algebra  $\mathfrak{g}$  of  $G = \mathbf{G}(\mathbb{R})$  in such a way that  $\|[u, v]\| \leq \|u\| \|v\|$  for all  $u, v \in \mathfrak{g}$ . We fix also an orthonormal basis for  $\mathfrak{g}$  w.r.t.  $\|\cdot\|$ .

We fix a semisimple subgroup  $H \subset G$  that satisfies the conditions indicated in §1.2.

Fix also an embedding  $\rho : \mathbf{G} \rightarrow \mathrm{GL}_N$ . We shall assume that the adjoint representation of  $\mathbf{G}$  occurs as an irreducible subrepresentation of that defined by  $\rho$ .

We take a *congruence lattice*  $\Gamma \leq G$ . That  $\Gamma$  is congruence means that  $\rho(\Gamma)$  contains a congruence subgroup of  $\mathrm{GL}_N(\mathbb{Z}) \cap \rho(G)$ .

Set  $X = \Gamma \backslash G$ .

The Lie algebra  $\mathfrak{g}$  possesses a natural  $\mathbb{Q}$ -structure. We choose a rational  $\Gamma$ -stable lattice  $\mathfrak{g}_{\mathbb{Z}} \subset \mathfrak{g}$  satisfying  $[\mathfrak{g}_{\mathbb{Z}}, \mathfrak{g}_{\mathbb{Z}}] \subset \mathfrak{g}_{\mathbb{Z}}$ . This is always possible, assuming only that  $\Gamma$  is arithmetic.

We fix throughout this section, and, indeed, throughout the entire paper, a homomorphism  $\phi : \mathrm{SL}_2(\mathbb{R}) \rightarrow H$  that projects nontrivially to each simple factor of  $H^+$ . This determines a unipotent one-parameter subgroup:

$$(3.1) \quad u(t) := \phi \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}.$$

The adjoint representation of this  $\mathrm{SL}_2(\mathbb{R})$ , i.e.  $\mathrm{Ad} \circ \phi$ , decomposes  $\mathfrak{g}$  into a direct sum of irreducible representations. Let  $\mathfrak{g}_0$  be the sum of all the highest weight spaces in all these irreducible representations, with respect to the diagonal torus of  $\mathrm{SL}_2(\mathbb{R})$  (equivalently: the  $\{u(t)\}$ -fixed subspace) and let  $\mathfrak{g}_1$  be the sum of all remaining weight spaces. Thus

$$(3.2) \quad \mathfrak{g} = \mathfrak{g}_0 \oplus \mathfrak{g}_1.$$

For  $r \in \mathfrak{g}$ , we shall write  $r = r_0 + r_1$  according to (3.2).

**3.1. Concerning constants and their implicit dependencies.** The notation  $\kappa, \kappa_1, \dots$  will always denote positive constants that depend *only on the isomorphism class of  $(H, G)$* . Because, in the circumstance of the main Theorem, there exist only finitely many possibilities for this isomorphism class if one fixes  $\dim(G)$  (cf. Lemma A.1.1), it is equivalent to say that these constants depend only on  $\dim(G)$ . Moreover, these constants are all indexed, and they come with hyperlinks to the point where they are defined. We hope the latter feature will be useful for readers using suitable viewing software.

An important note is that, because there are only finitely many intermediate subgroups  $S$  between  $H$  and  $G$ , any constant that depends only on  $S$  or just on its isomorphism class also can be chosen so that it depends only on  $H$ , and  $G$ . We shall use this observation several times without explicit mention.

The notation  $\iota, \iota_1, \dots$  will denote positive constants that depend only<sup>6</sup> on  $\mathbf{G}, H, \rho$ , the norm  $\|\cdot\|$ , and on the given orthogonal basis of  $\mathfrak{g}$ .

The notation  $c_1, c_2, \dots$  will denote positive constants that may depend on  $\mathbf{G}, H, \rho, \|\cdot\|$ , the lattice  $\Gamma$ , and on the lattice  $\mathfrak{g}_{\mathbb{Z}} \subset \mathfrak{g}$  discussed in §3.6. The numbering of these constants resets by subsection.

If either of these constants depends additionally on other parameters we make this explicit by writing them in parenthesis, e.g.  $\iota_2(d)$  is a constant depending on  $\mathbf{G}, H, \rho$ , the norm  $\|\cdot\|$ , the orthogonal basis of  $\mathfrak{g}$ , and the parameter  $d$ .

---

<sup>6</sup>In particular, constants of the form  $\iota_*$  are allowed to depend on the  $\mathbb{Q}$ -structure on  $\mathbf{G}$ , by contrast with  $\kappa_*$ .

As a rule of thumb we shall strive to ensure that the *exponents* in our results depend only on  $G, H$ ; on the other hand, we shall not strive for such minimal dependency in other constants. For instance, in the statement of the Theorem, the exponent  $\delta$  depends only  $G, H$ ; whereas the constant  $V_0$  is permitted to depend on  $\Gamma$ .

**3.2. The  $\ll$ -notation and the  $\star$ -notation.** So that the notation of this paper does not become overwhelming, we shall introduce certain notational conventions.

We shall use the expression “ $A \ll B$ ,” for two positive quantities  $A, B$ , to mean that  $A \leq c_3(d)B$ , notations as above. (It will happen in a large fraction of this paper that a parameter  $d$  will be present, measuring the index of a suitable Sobolev norm. Therefore, we make the convention that implicit constants are allowed to depend on the symbol “ $d$ ”, or whatever its value in the present context is. If no symbol  $d$  is present,  $A \ll B$  simply means  $A \leq c_4B$ . In §3.7 we give an example to clarify this notation.)

We shall use  $A \asymp B$  to mean that both  $A \ll B$  and  $B \ll A$ .

Suppose  $A$  is a quantity taking values in  $(0, \infty)$ . We shall write  $A_\uparrow$  for any quantity  $f(A)$  that is defined for sufficiently large  $A$  in a fashion that depends only on  $G, H$ , and so that  $f(A) \rightarrow \infty$  as  $A \rightarrow \infty$ . We write  $A_\downarrow$  for any quantity  $g(A)$  defined for sufficiently small  $A$  in a fashion that depends only on  $G, H$  and so that  $g(A) \rightarrow 0$  as  $A \rightarrow 0$ .

For instance, the function  $A^{\dim(H)} - \dim(G)A^{\dim(H)-1}$  could be abbreviated as  $A_\uparrow$ .

We write  $B = A^\star$  if  $B = c_5A^{\kappa_2}$ . We write  $B = A^{-\star}$  if  $B = c_6A^{-\kappa_2}$ . In a similar fashion, we define  $B \leq A^\star$ ,  $B \geq A^\star$ , etc.

For instance, the function  $\text{vol}(\Gamma \backslash G)A^{\dim G}$  could be abbreviated as  $A^\star$ .

**3.3. Metrics, measures, Lie algebra.** The Euclidean norm on  $\mathfrak{g}$  defines a left-invariant Riemannian metric on  $G$ , which descends to a metric on  $X = \Gamma \backslash G$ .

For  $g_1, g_2 \in G$ , we write  $g_1 \stackrel{\epsilon}{\sim} g_2$  if the distance between  $g_1$  and  $g_2$  is  $\leq \epsilon$ . Similarly, we use the same notation  $x_1 \stackrel{\epsilon}{\sim} x_2$  for  $x_1, x_2 \in X$ . We say a finite subset of a metric space is  $\delta$ -separated if the points all are at mutual distance  $\geq \delta$ .

The Riemannian metric on  $G$  also gives a Haar measure on any subgroup of  $G$ , in particular, all the intermediate subgroups between  $H$  and  $G$ . We shall denote these measures by  $\text{dvol}$ . If  $Q$  is a subgroup of  $G$ , we may use this Haar measure to speak of the *volume*  $\text{vol}(x_0Q)$  of any closed  $Q$ -orbit on  $X$ . By contrast, we shall use either the letters  $\mu$  or  $\nu$  to denote the  $Q$ -invariant probability measure on  $x_0Q$ .

We set  $\|g\| := \max_{ij} (|\rho(g)|_{ij}, |\rho(g^{-1})|_{ij})$ .

We note that, for some constant  $\iota_3$

$$(3.3) \quad \|g^{-1}\| = \|g\|, \quad \|g_1g_2\| \leq N\|g_1\|\|g_2\|, \quad \|\text{Ad}(g)\|_{\text{op}} \leq \iota_3\|g\|$$

Here  $\|\cdot\|_{\text{op}}$  denotes the operator norm w.r.t. the chosen Euclidean norm on  $\mathfrak{g}$  and the last inequality follows since we assume that the adjoint representation occurs in  $\rho$ . A consequence of (3.3) is

$$(3.4) \quad d(xg, yg) \leq \iota_3 \|g\| d(x, y)$$

either for  $x, y \in X$  or  $x, y \in G$ .

**3.4. Intermediate subgroups.** For each intermediate subgroup  $H \subset S \subset G$ , we let  $S^0$  be the connected component of the identity in  $S$ , and let  $\tilde{S}$  be the normalizer of the Lie algebra  $\mathfrak{s}$ ; thus  $S^0 \subset S \subset \tilde{S}$ ; moreover, the index  $[\tilde{S} : S^0]$  is finite by virtue of the assumption that  $\mathfrak{h}$  has trivial centralizer. Also,  $\tilde{S}$  is “a real algebraic group”, i.e. consists of the real points of an algebraic subgroup of  $\mathbf{G}$  defined over  $\mathbb{R}$ . The following will be proved in §A.

**3.4.1. Lemma.** *Suppose  $H \subset G$  is semisimple and that  $\mathfrak{h}$  has trivial centralizer in  $\mathfrak{g}$ . Then  $H$  is not contained in any proper parabolic subgroup of  $G$ . Moreover, there are only finitely many intermediate subgroups  $H \subset S \subset G$ . Each such  $S$  is semisimple without compact factors.*

Although we shall not explicitly introduce notation for it, it is convenient to regard as fixed a choice of maximal compact subgroup of  $S$ , for every intermediate subgroup  $H \subset S \subset G$ .

**3.5. Balls in intermediate subgroups.** For each connected intermediate subgroup  $S \supset H$ , we shall define a certain increasing family of balls  $B_S(T)$  (sometimes denoted  $B_S^T$ ). These balls have the following properties:

- (1)  $B_S(T) \subset B_S(T')$  if  $T \leq T'$ ;
- (2)  $B_S(T) \subset \{s \in S : \|s\| \leq T\}$ .
- (3) (a) If  $\Omega \subset S$  is compact, there exists  $c = c_7(\Omega)$  so that  $\Omega.B_S(T).\Omega \subset B_S(cT)$ ;
- (b) As  $\Omega \downarrow 1$ , we may take  $c \rightarrow 1$ .
- (4) The volume  $B_S(T)$  grows as a positive power of  $T$  (up to logarithms), see (6.12) for the precise statement.

The construction is slightly elaborate; taking simply  $B_S(T) = \{s \in S : \|s\| \leq T\}$  is fine when  $S$  is almost simple, but gives rise to “hyperbola-like” balls in the semisimple case, so a slight modification is needed. The reader need not pay too much attention to the precise details of the construction.

Fix for each such  $S$  an isogeny

$$(3.5) \quad \Phi : \prod_{i=1}^I S_i \rightarrow S$$

from a product of connected, almost-simple groups onto  $S$ . We put, for  $S$  connected,

$$B_S(T) := \Phi\{(s_1, \dots, s_I) : \|\Phi(s_i)\| \leq N^{-1}T^{1/I} \text{ for } i = 1, \dots, I\}.$$

It would be possible to define corresponding sets in the disconnected case, but we do not need this.

**3.6. Noncompactness.** For  $x \in X$  we set:

$$\begin{aligned} \text{ht}(x) &:= \sup\{\|\text{Ad}(g^{-1})v\|^{-1} : \Gamma g = x, v \in \mathfrak{g}_{\mathbb{Z}} \setminus \{0\}\} \\ \mathfrak{S}(R) &:= \{x \in X : \text{ht}(x) \leq R\}. \end{aligned}$$

In words  $\text{ht}(x)^{-1}$  is the size of the smallest nonzero vector in the lattice  $\text{Ad}(g^{-1})\mathfrak{g}_{\mathbb{Z}} \subset \mathfrak{g}$  corresponding to  $x = \Gamma g$ . Note that we have  $\text{ht}(x) \gg 1$  for all  $x \in X$ .

As follows from reduction theory the set  $\mathfrak{S}(R)$  is a compact subset of  $X$ . Indeed, every  $x \in \mathfrak{S}(R)$  may be expressed as  $x = \Gamma g$ , where  $\text{Ad}(g)$  has operator norm  $\|\text{Ad}(g)\|_{\text{op}} \ll R$  w.r.t. the norm on  $\mathfrak{g}$ .

We note that

$$(3.6) \quad \text{ht}(xg) \leq \iota_3 \|g\| \text{ht}(x).$$

Moreover, there are constants  $\iota_4$  and  $\kappa_3$  so that

$$(3.7) \quad g \mapsto xg \text{ from } G \text{ to } X \text{ is injective for } d(g, 1) \leq \iota_4 \text{ht}(x)^{-\kappa_3}.$$

That statement follows by considering a basis for the lattice  $\text{Ad}(g^{-1})\mathfrak{g}_{\mathbb{Z}}$  corresponding to  $x = \Gamma g$ . We shall refer to  $\iota_4 \text{ht}(x)^{-\kappa_3}$  as the *injectivity radius* at  $x$ .

We shall require the following lemma, which relies on the mentioned linearization technique [27] and on our technical assumption in §1.2 that the centralizer of  $\mathfrak{h}$  is trivial (in the form of Lemma 3.4.1). The proof is given in Appendix B.

**3.6.1. Lemma.** (1) *There exists  $R_1 \geq 1$  so that any  $H$ -orbit on  $X$  intersects  $\mathfrak{S}(R_1)$ ;*

(2) *There are constants  $c_8$  and  $\kappa_4$  so that, for any  $H$ -invariant measure  $\mu$ ,*

$$\mu(\{x \notin \mathfrak{S}(R)\}) \leq c_8 R^{-\kappa_4}$$

In particular, we may choose  $R_0$  so large that, for any  $H$ -invariant measure  $\mu$ , we have  $\mu(X \setminus \mathfrak{S}(R_0/2)) \leq 10^{-11}$ . The set  $\mathfrak{S}(R_0)$  will occur throughout our paper as a convenient choice of compact. We shall refer to it as  $X_{\text{cpct}}$ .

**3.7. Sobolev norms.** In dealing with analytic questions on  $X = \Gamma \backslash G$ , we shall make systematic use of a certain family of Sobolev norms. While there is overhead in developing their basic properties, they offer a clean formalism for analysis on homogeneous spaces.

Recall that we have chosen an orthonormal basis for  $\mathfrak{g}$ . This defines for every  $d \geq 0$  a system of  $L^2$ -Sobolev norms  $\mathcal{S}_d$  on  $C_c^\infty(X)$ :

$$(3.8) \quad \mathcal{S}_d(f)^2 := \sum_{\mathcal{D}} \|\text{ht}(x)^d \mathcal{D}f\|_{L^2}^2,$$

where the sum is taken over  $\mathcal{D} \in U(\mathfrak{g})$ , the universal enveloping algebra of  $\mathfrak{g}$ , which are monomials in the chosen basis of degree  $\leq d$ .

Although defined **for all**  $d$ , we use only for  $d$  sufficiently large that  $\mathcal{S}_d$  majorizes pointwise values of derivatives; see important comment below.

We note that  $\mathcal{S}_d$  defines a *Hermitian* norm on  $C_c^\infty(X)$ . It has the following basic properties: there exists some  $\kappa_5$ , which we may assume to be larger than  $\dim(G)$  for technical convenience later, so that

$$(3.9) \quad \|f\|_{L^\infty} \ll_d \mathcal{S}_d(f), \text{ for } d \geq \kappa_5 \text{ and } f \in C_c^\infty(X)$$

and for every  $d$  there exists<sup>7</sup> an integer  $d' > d$

$$(3.10) \quad \text{Trace}(\mathcal{S}_d^2 | \mathcal{S}_{d'}^2) < \infty$$

This means: if we complete  $C_c^\infty(X)$  in the norm defined by  $\mathcal{S}_{d'}$  to obtain a Hilbert space  $V_{d'}$ , the form  $\mathcal{S}_d(f)^2 = \langle A_d f, f \rangle$  is defined by a positive, compact operator  $A_d : V_{d'} \rightarrow V_{d'}$  with finite trace. In alternate terms, there exists an orthonormal basis  $\{e_i\}$  w.r.t.  $\mathcal{S}_{d'}$  so that  $\sum_i \mathcal{S}_d(e_i)^2 < \infty$ .

To establish (3.9) and (3.10), we use (3.7) and reduce to  $\mathbb{R}^{\dim G}$  via coordinate patches. We give details in §5.

We note that there is  $\kappa_6$  and  $\iota_5(d)$  (where the latter depends also on  $d$ ) so that

$$(3.11) \quad \mathcal{S}_d(g.f) \leq \iota_5(d) \|g\|^{\kappa_6 d} \mathcal{S}_d(f) \text{ for any } f \in C_c^\infty(X) \text{ and } g \in G$$

Here  $g \in G$  acts on a function  $f$  defined on  $X$  by  $(g.f)(x) = f(xg)$ . To see this, combine (3.3), (3.6), (3.8).

If  $f_1, f_2 \in C_c^\infty(X)$ , then,

$$(3.12) \quad \mathcal{S}_d(f_1 f_2) \ll_d \mathcal{S}_{d+\kappa_5}(f_1) \mathcal{S}_{d+\kappa_5}(f_2),$$

as may be deduced from the definition and (3.9).

Once  $d \geq \kappa_5 + 1$  we have

$$(3.13) \quad \|f - g.f\|_{L^\infty} \ll_d d(e, g) \mathcal{S}_d(f)$$

This follows from (3.9) applied to the partial derivatives of  $f$  and by integrating the directional derivative of  $f$  along a geodesic curve connecting  $x$  and  $xg$ .

**3.8. Notational conventions concerning Sobolev norms.** We observe two very important notational conventions concerning Sobolev norms.

Throughout this paper, when we write  $\mathcal{S}_d$ , we *always* assume that  $d \geq \kappa_5 + 1$ .

Moreover, in any statement that makes reference to a Sobolev norm  $\mathcal{S}_d$ , the implicit constants in the symbol  $\ll$  are permitted to depend on  $d$ .

As an example of these conventions: (3.12) could be legitimately abbreviated by  $\mathcal{S}_d(f_1 f_2) \ll \mathcal{S}_{d+\kappa_5}(f_1) \mathcal{S}_{d+\kappa_5}(f_2)$ .

---

<sup>7</sup>In this discussion of traces, we draw heavily on the work [2] of Bernstein and Reznikov.

**3.9.  $L^2$ -spaces.** If  $\nu$  is a measure, we denote by  $L^2(\nu)$  the associated  $L^2$ -Hilbert space and by  $L_0^2(\nu)$  the orthogonal complement of the constant function.

In some cases, it will be more natural to use  $L_0^2$  to denote the orthogonal complement of *locally* constant functions: for instance, if the support of  $\nu$  is disconnected. We will always indicate clearly when this is the case.

**3.10. Almost invariant measures.** For any  $g \in G$  and any measure  $\nu$  on  $X$ , we denote by  $\nu^g$  the translated measure that is defined via  $\nu^g(f) = \int f(xg)d\nu(x)$  for  $f \in C_c(X)$  or equivalently  $\nu^g(B) = \nu(Bg^{-1})$  for any measurable  $B \subset X$ .

In what follows, we shall define several notions of a measure being *almost invariant*. More precisely, these notions will be *almost invariant w.r.t. to a Sobolev norm  $\mathcal{S} = \mathcal{S}_d$* . We will sometimes omit reference to the Sobolev norm if it is clear from context.

We say that  $\mu$  is  $\epsilon$ -almost invariant under  $g \in G$  if

$$|\mu^g(f) - \mu(f)| \leq \epsilon \mathcal{S}(f), \quad f \in C_c^\infty(X)$$

We say that it is  $\epsilon$ -almost invariant under a subgroup  $Q$ , if it is  $\epsilon$ -almost invariant under every  $q \in Q$  with  $\|q\| \leq 2$ . We say that it is  $\epsilon$ -almost invariant under  $Z \in \mathfrak{g}$  if it is  $\epsilon$ -almost invariant under every  $\exp(tZ)$  with  $|t| \leq 2$ .

These notions satisfy all the expected properties: e.g. if a measure is  $\epsilon$ -almost invariant under  $Z_1, Z_2 \in \mathfrak{g}$  with  $\|Z_1\|, \|Z_2\| \leq 2$ , it is  $\epsilon^*$ -almost invariant under  $Z_1 + Z_2$ .

A list of such properties is given, with proof, in §8.

**3.11. Spectral gap.** There exists an integer  $p_G$ , depending only on  $G$ , with the following property.

Let  $H \subset S \subset G$  be as before, and let  $x_0 \in X$  be so that  $x_0S$  is a closed connected orbit. Let  $\nu$  be the  $S$ -invariant probability measure on  $x_0S$ , and  $L_0^2(\nu)$  the orthogonal complement of the constant function. Then  $L_0^2(\nu)^{\otimes(p_G-1)}$  is *tempered* as an  $S$ -representation; for a review of this notion, see §6.2. This applies, in particular, to the measure  $\mu$  as in the statement of Theorem 1.3. We discuss in §6 the origins of that statement and its consequences that we will need.

#### 4. OUTLINE OF THE PROOF OF THEOREM 1.3.

This section is intended as a summary of the paper. We give a detailed outline of the proof of Theorem 1.3, giving references and comparisons to the non-effective proof in §2. The actual proof of the Theorem is given, along the lines indicated here, in §16.

While this outline is intended to be a reasonable résumé of the proof, we nonetheless have not indicated all the technicalities that are involved in the complete proof.

We remark that the presentation of this section is designed so as to match §2. The order in which results are discussed here, therefore, does not always match the order in the body of the paper.<sup>8</sup>

**4.1. Setup.** Let, for the whole section,  $G$ ,  $H$ ,  $\Gamma$ ,  $X = \Gamma \backslash G$ , and  $\mu$  be as in the statement of Theorem 1.3. Let  $S \supset H$  be a closed subgroup. As in the proof in §2 we let  $\mathfrak{r}$  be an  $\text{Ad}(H)$ -invariant complement of the Lie algebra  $\mathfrak{s}$  of  $S$  in  $\mathfrak{g}$ . Recall that we have fixed a one-parameter unipotent subgroup  $u(t)$  of  $H$  that projects nontrivially to each simple factor.

We shall assume that  $\mu$  is almost invariant under  $S$ , and will demonstrate that *either* it is almost invariant under a larger subgroup; or  $\mu$  is supported on a closed  $S$ -orbit of small volume.

**4.2. Outline of the steps.** As discussed in §1.5 one can regard the proof in three steps:

- (1) *Ergodic theorem.* The assertion that “most points are generic” is used in §2.4 in the ineffective proof, and appears as Proposition A (§4.3).
- (2) *Nearby generic points give additional invariance.* A form of this statement appears in §2.5 in the ineffective proof; an effective form will be given here in Proposition B (§4.4).
- (3) *Dichotomy:* If we cannot find two nearby generic points, as above, then necessarily  $\mu$  was supported on a closed  $S$ -orbit of small volume. This appears in §2.6 of the ineffective proof, and an effective form will appear here as Proposition C (§4.5). The proof of this is the most involved argument of the paper, involving a closing result for actions of semisimple groups: an “almost-closed” orbit is near a closed orbit.

The conclusion of the non-effective proof, §2.7, makes use of some seemingly trivial principles: If  $\mu$  is invariant under  $S$  and under  $g^* \notin S$ , then it is invariant by a strictly larger subgroup  $S_* \supset S$ ; and, if  $\mu(xS) > 0$ , then  $\mu$  is the  $S$ -invariant measure on the closed orbit  $xS$ . To get the main Theorem from Propositions A, B and C requires that these facts be made effective. We discuss these issues in §4.6 and conclude the sketch of proof of the theorem in §4.7.

**4.3. Effective ergodic theorem (§2.4).** Fixing a suitable big integer  $M$  – the choice depends only on  $G$  – we say that a point  $x \in X$  is  $[T_0, T_1]$ -generic w.r.t.  $\mu$  if the natural probability measure on  $\{u(t)x\}_{T^M \leq t \leq (T+1)^M}$  approximates  $\mu$  to within an error of about  $T^{-1}$ , whenever  $T \in [T_0, T_1]$  is an integer. The choice of  $M$  and a precise formulation of this is to be found in the discussion around (9.1).

---

<sup>8</sup>The primary reason for this is that a certain group of results, which are used at different points in the argument, have their roots in the polynomial divergence properties of unipotent flows. Therefore, we treat them together in the body of the paper.

**Proposition A.** (*Proposition 9.3*) Let  $H \subset S \subset G$ ,  $S$  connected. Suppose that  $\mu$  is  $\epsilon$ -almost invariant under  $S$  w.r.t.  $\mathcal{S}_d$ , for  $d \geq \kappa_5 + 1$ . Then there exists  $\beta \in (0, 1/2)$ ,  $d' > d$ , depending only on  $G$ ,  $H$ , and  $d$ , so that:

Whenever  $R \leq \epsilon^{-\beta}$  and  $T_0 > 0$ , the fraction of points  $(x, s) \in X \times B_S(R)$  (w.r.t.  $\mu$  resp. the Haar measure on  $S$ ) for which  $x.s$  is not  $[T_0, \epsilon^{-\beta}]$ -generic w.r.t.  $\mathcal{S}_{d'}$  is  $\ll_d T_0^{-1}$ .

This assertion is the effective replacement of the discussion of §2.4.

Let us contrast more carefully the above statement with that of §2.4. The Birkhoff ergodic theorem says, roughly speaking, that the measure of  $[T_0, \infty)$ -generic points approaches 1, as  $T_0 \rightarrow \infty$  (but doesn't give an error either in the rate of genericity nor in the measure of the set). §2.4 isolates a large subset  $E$  of such points. In the context of §2.4, because the measure  $\mu$  was *exactly*  $S$ -invariant, the set of pairs  $(x \in X, s \in S)$  so that  $xs \in E$  has large measure.

In the almost-invariant context, one obtains only that the set of pairs  $(x \in X, s \in S)$  so that  $xs$  is *very close* to  $E$  has large measure. In particular, for such pairs  $(x, s)$ , the point  $xs$  is not  $[T_0, \infty)$ -generic, but only  $[T_0, T_1]$ -generic, where the size of  $T_1$  depends on the strength of the notion of *very close*.

This discussion accounts for the difference in formulation between Proposition A and §2.4.

Proposition A is proved using the quantitative information about decay of correlations provided by the spectral gap. It also makes use of the trace estimate in (3.10).

**4.4. Nearby generic points give additional invariance (§2.5).** We refer to (3.2) for the notation  $r_1$ , for  $r \in \mathfrak{r}$ .

**Proposition B.** (*Proposition 10.2*) Let  $d \geq \kappa_5 + 1$ . There exists constants  $\kappa_7 > 0$  and  $\kappa_8 > \kappa_9 > 0$  so that:

Suppose  $\mu_1, \mu_2$  are  $H$ -invariant measures, that  $x_1, x_2 \in X$  satisfy  $x_2 = x_1 \exp(r)$  for some nonzero  $r \in \mathfrak{r}$ , and that  $x_i$  is  $[\|r\|^{-\kappa_8}, \|r_1\|^{-\kappa_9}]$ -generic w.r.t.  $\mu_i$  and a Sobolev norm  $\mathcal{S}_d$  (for  $i = 1, 2$ ). Then there is a polynomial  $q : \mathbb{R} \rightarrow \mathfrak{r}$  of degree  $\leq \dim(\mathfrak{g})$ , so that:

$$\left| \mu_2^{\exp q(s)}(f) - \mu_1(f) \right| \ll_d \|r_1\|^{\kappa_7} \mathcal{S}_d(f), 1 \leq s \leq 2^{1/M}$$

If  $r_1 \neq 0$ , then  $\max_{s \in [0, 2]} \|q(s)\|$  can be set to be one.<sup>9</sup> Moreover, if  $\mu_1 = \mu_2$  is  $\epsilon$ -almost invariant under  $S$ , then  $\mu_1$  is  $\ll_d \max(\epsilon, \|r\|^{\kappa_7})^{1/2}$ -almost invariant under some  $Z \in \mathfrak{r}$  with  $\|Z\| = 1$ .

This Proposition, based on polynomial divergence, constitutes an effective version of the discussion from §2.5. More specifically, when applied with

<sup>9</sup>However, we will need to set this maximum equal to some constant that depends only on  $G$  and  $H$ .

$\mu_1 = \mu_2$ , it yields an effective version of Lemma 2.5.1.<sup>10</sup> The generalization to two distinct measures  $\mu_1 \neq \mu_2$  requires no effort, and is technically convenient for certain other applications (e.g., when studying two distinct closed  $S$ -orbits).

The proof of this Proposition is an obvious quantification of §2.5. The Proposition and related ideas have several corollaries that will be used later in the proof. In particular, it implies a quantitative version of the isolation of distinct closed  $S$ -orbits.

#### 4.5. Dichotomy (§2.6).

**Proposition C.** (*Proposition 14.2*) *Given a connected intermediate subgroup  $H \subseteq S \subseteq G$ ,  $\zeta \in (0, 1)$ , and  $d_S \geq 1$ , there exists  $\xi$  and  $d'$  depending only on  $d_S$ ,  $\zeta$ ,  $G$ ,  $H$ , and  $\epsilon_0 \ll_{d_S} 1$ , so that:*

*Suppose for some  $\epsilon \leq \epsilon_0$  that*

- (1)  *$\mu$  is  $\epsilon$ -almost invariant under  $S$ , with respect to the Sobolev norm  $\mathcal{S}_{d_S}$ , and that*
- (2)  *$\mu(xS) = 0$  for all closed  $S$ -orbits of volume  $\leq \epsilon^{-\zeta}$ .*

*Then there exists  $x_1, x_2$  so that  $x_2 = x_1 \exp(r)$ ,  $r \in \mathfrak{r}$ ,  $\|r\| \leq \epsilon^\xi$ , and  $x_1, x_2$  are both  $[\|r\|^{-\kappa_8}, \|r_1\|^{-\kappa_9}]$ -generic w.r.t.  $\mathcal{S}_{d'}$ .*

This Proposition is an effective analog of §2.6, in particular, an effective analog of Lemma 2.6.1. However, the proof is considerably more involved. It necessitates, in particular, an “effective closing result”: the assertion that if an orbit  $xS$  is “almost closed”, there exists  $x'$  near  $x$  that is closed.

We devote most of the present subsection to explaining the proof of Proposition C. We begin by enunciating two results which are, in spirit, quite close to Proposition B (= Proposition 10.2); for, like that Proposition, they are based in concept on the polynomial divergence properties of unipotent flows. For that reason, in the text, they are proved immediately after Proposition 10.2.

**4.5.1. Lemma.** (*Lemma 10.3.1*) (*Quantitative isolation of periodic orbits for semisimple groups*) *There are constants  $\kappa_{10}$  and  $\kappa_{11}$  with the following property.*

*Suppose  $H \subset S \subset G$ . Let  $x_1, x_2 \in X$  be so that  $x_i S$  are closed orbits with volume  $\leq V$ . Then either:*

- (1)  *$x_1$  and  $x_2$  are on the local  $S$ -orbit, i.e. there exists some  $s \in S$  with  $d(s, 1) \leq 1$  and  $x_2 = x_1 s$ , or*
- (2)  *$d(x_1, x_2) \gg \min(\text{ht}(x_1), \text{ht}(x_2))^{-\kappa_{10}} V^{-\kappa_{11}}$ .*

The Lemma in combination with Lemma 3.6.1 implies, in particular, that the *total* number of closed  $S$ -orbits with volume  $\leq V$  is bounded by a polynomial in  $V$ .

<sup>10</sup>See §1.7 for a discussion of the origins of that argument.

**Proposition C1.** (*Proposition 11.1*) *Let  $S \supset H$ . There exists some  $V_0$  depending on  $\Gamma$ ,  $G$ , and  $H$  and some  $\kappa_{12} > 0$  with the following property. Let  $V \geq V_0$  and suppose  $\mu(Y) = 0$  if  $Y$  is any closed  $S$ -orbit of volume  $\leq V$ . Then:*

$$(4.1) \quad \mu(\{x \in X : \text{there exists } x' \stackrel{V^{-\kappa_{12}}}{\sim} x \text{ with } \text{vol}(x'S) \leq V\}) \leq 1/2$$

This Proposition is again based on polynomial divergence. For each closed orbit  $x'S$  as in (4.1), the  $\mu$ -mass of a  $V^{-\kappa_{12}}$ -neighbourhood of  $x'S$  is quite small. The idea of this is that one can linearize the flow in the neighbourhood of  $x'S$  and thereby understand it completely. Taken in combination with Lemma 4.5.1 (10.3.1), this yields the claimed result.

**Proposition C2.** (*Proposition 13.1*) *Let  $H \subset S \subset G$ , with  $S$  connected. Let  $\delta \leq 1 \leq N$ . There exists some  $T_0 = T_0(\Gamma, G, H, N)$  with the following property:*

*Let  $T \geq T_0$  and let  $v = \text{vol } B_S(T)$ . Suppose that  $\{s_1, \dots, s_k\} \subset B_S(T)$  is  $\frac{1}{10}$ -separated, that  $k \geq v^{1-\delta}$ , and that there exists  $x \in X_{\text{cpct}}$  so that  $xs_i \stackrel{T^{-N}}{\sim} xs_j$ . Then there is  $x' \stackrel{T^{-N}}{\sim} x$  so that  $x'S$  is a closed orbit of volume  $\leq T^{\delta \downarrow}$ .*

This result constitutes an effective “closing Lemma” for actions of semisimple groups on homogeneous spaces.

*Idea of proof.* Choosing a representative  $g_0 \in G$  for  $x$ , the fact that  $xs_i$  is close to  $xs_j$  means that there is an element  $\gamma \in \Gamma$  that moves  $g_0s_i$  near to  $g_0s_j$ . Thereby, one constructs a whole collection of elements of  $\Gamma$  near to  $g_0Sg_0^{-1}$ .

We first establish that all these elements must themselves lie on a different conjugate  $g'_0Sg_0'^{-1}$ . Here  $g'_0 \in G$  is very close to  $g_0$ . This uses, in particular, the arithmetic nature of  $\Gamma$ . A baby version of this argument is the fact that, if three points in  $\mathbb{Z}^2 \cap [-N, N]^2$  all lie within  $\frac{1}{100N}$  of a certain line in the plane, then they must *exactly* lie on a nearby line.

Next, we show that the existence of so many elements in  $\Gamma$  that lie inside  $g'_0Sg_0'^{-1}$  force the subgroup  $g'_0Sg_0'^{-1} \cap \Gamma$  to be a lattice. This step uses a spectral gap. A simple statement with the same general flavor as what we need is the following: Suppose we are given a discrete subgroup  $\Lambda \subset \text{SL}_3(\mathbb{R})$  so that

$$(4.2) \quad \lim_{T \rightarrow \infty} \frac{\#\{m \in \Lambda, \|m\| := \sqrt{\text{Trace}(m \cdot m^t)} \leq T\}}{T^{5.9}} = \infty$$

Then  $\Lambda$  is a lattice. (For comparison, the asymptotic for  $\text{SL}_3(\mathbb{Z})$  is  $T^6$ .)

This statement may be derived from Property (T). (The fact that  $\Lambda$  has many elements translates to the existence of a subrepresentation of  $L^2(\Lambda \backslash \text{SL}_3(\mathbb{R}))$  near to the trivial representation.) The precise form of what we use is somewhat different and is stated in Proposition 12.2, part (2).

Therefore, we have established that  $x' := \Gamma g'_0$  is close to  $x = \Gamma g_0$ , and  $x'S$  is a closed  $S$ -orbit. A more detailed study shows that the volume of  $x'S$  is also small.  $\square$

Now, we sketch the proof of Proposition C.

By Proposition A, for “most”  $x \in X$ , the measure of  $\mathcal{B}_x := \{s \in S : xs \text{ generic}\}$  is “large.” (Here “generic” means the notion specified in the statement of Proposition C.)

More is true: let  $O$  be a small neighbourhood of the identity in  $S$ . Given  $x \in X$ , let  $\mathcal{B}'_x$  consist of  $s \in S$  so that  $\text{vol}(sO \cap \mathcal{B}_x) > 0.99 \text{vol}(O)$ . By Fubini’s theorem, the volume of  $\mathcal{B}'_x$  is large for “most”  $x$ . (Compare the usage of Fubini’s theorem in §2.6.)

Supposing, for simplicity, that  $X$  were compact, we cut  $X$  into boxes of size  $\epsilon^*$ . Let us observe that if one such  $\epsilon^*$  box contained a large fraction of the volume of  $x.\mathcal{B}'_x$ , then Proposition C2 would show that there is  $x'$ , very near to  $x$ , so that  $x'S$  is closed. Proposition C1 shows that this cannot happen, at least for “most”  $x$ . Therefore, Proposition C1 shows that, for at least one  $x \in X$ , no single  $\epsilon^*$ -box contains a large fraction of the volume of  $\mathcal{B}_x$ .

Consequently, for at least one  $x \in X$ , there exist two distinct but close  $\epsilon^*$ -boxes, each of which contain a point of the form  $xs, s \in \mathcal{B}'_x$ . It is easy to see – by “adjusting along the  $S$ -direction” – that this implies the conclusion of Proposition C.

**4.6. Effective version of §2.7.** In this section, we discuss some of the results which effectivize the rather trivial-seeming principles used implicitly in §2.7. After this, we present the proof of Theorem 1.3.

**Proposition D.** (*Proposition 8.1*) *Fix a Sobolev norm  $\mathcal{S}_d$ ; all notions of almost invariance are taken with respect to this. Suppose that  $\mu$  is  $\epsilon$ -almost invariant under a subgroup  $S$  and also under  $Z \in \mathfrak{g}$  such that  $Z \perp \mathfrak{s}$  and  $\|Z\| = 1$  where  $\mathfrak{s}$  is the Lie algebra of  $S$ .*

*Then there is a constant  $\kappa_{13}$  so that  $\mu$  is also  $c_1(d)\epsilon^{\kappa_{13}}$ -almost invariant under some subgroup  $S_*$  with  $\dim(S_*) > \dim(S)$ . If  $H \subseteq S$ , we may also assume that  $H \subseteq S_*$ .*

This is proved in §8, and effectivizes the statement “If  $\mu$  is invariant under  $S$  and under  $g^* \notin S$ , then it is invariant by a strictly larger subgroup  $S_* \supset S$ ” used in §2.7. The difficulty may be seen in the following case: let  $S_1$  be an intermediate subgroup, and suppose that  $Z$  lies *almost* in the Lie algebra of  $S_1$ . Then, in the statement of the proposition, one wants to take  $S_* = S_1$  rather than the group  $S_2$  generated by  $H$  and  $\exp(Z)$ . Indeed, because  $Z$  lies almost in  $S_1$ , the subgroup  $H$  and the element  $\exp(Z)$  generate  $S_2$  very “inefficiently.”

We deduce it from the corresponding Lie algebra statement. Given  $T = \{t_1, \dots\} \subset \mathfrak{g}$ , we set  $T^{(k)}$  to be the set of all possible iterated Lie brackets of the  $t_i$ s of depth  $\leq k$ .

**Proposition E.** (Proposition 7.1) *There exists an integer  $k$  and some  $c > 0$  depending only on  $G$  with the following property. For any orthonormal subset  $T = \{t_1, \dots\}$  of  $\mathfrak{g}$ , and  $0 < \delta < 1$  there exists a subalgebra  $\mathfrak{w} \subset \mathfrak{g}$  with orthonormal basis  $w_1, w_2, \dots$  satisfying:*

- (1) *For each  $w_i$ , there exists a linear combination  $w'_i = \sum_{t \in T^{(k)}} c_t t$ , with  $c_t \in \mathbb{R}$  satisfying  $|c_t| \leq \delta^{-k}$ , so that  $\|w_i - w'_i\| \ll \delta^c$ .*
- (2) *Each  $t \in T$  is within  $\delta$  of  $\mathfrak{w}$  (i.e.  $\min_{w \in \mathfrak{w}} \|w - t\| \leq \delta$ ).*

*If the linear span  $\langle T \rangle$  of  $T$  contains a subalgebra  $\mathfrak{h}$ , then there exists  $k$  and  $c$  depending on the pair  $(\mathfrak{h}, \mathfrak{g})$  so that (1) and (2) hold and, in addition,  $\mathfrak{h} \subset \mathfrak{w}$ .*

In explicit terms,  $\mathfrak{w}$  is a Lie subalgebra that is efficiently generated by small perturbations of the elements of  $T$ . The proof of this is (at least intuitively) clear, the  $t_i$  span an “almost-subalgebra” in an efficient way, and because the space of subalgebras of  $\mathfrak{g}$  is compact, this almost-subalgebra is near a genuine algebra. We will use the Łojasiewicz inequality to carry that argument through in an effective way.

**Proposition F.** (Proposition 15.1) *Let  $x_0 \in X$  be so that  $x_0 S$  is a closed orbit of volume  $V$ , for some connected  $S \supset H$ . Suppose  $\mu$  is a probability measure on  $x_0 S$  that is  $\epsilon$ -invariant under  $S$  w.r.t. a Sobolev norm  $\mathcal{S}_d$ . Let  $\nu$  be the  $S$ -invariant probability measure on  $x_0 S$ .*

*Then there are  $\kappa_{14}$  and  $\kappa_{15}$  so that*

$$(4.3) \quad |\mu(f) - \nu(f)| \ll_d V^{\kappa_{14}} \epsilon^{\kappa_{15}/d} \mathcal{S}_d(f) \text{ for } f \in C_c^\infty(X).$$

*In particular, there are constants  $\kappa_{16}, \kappa_{17} > 0$  such that if  $V \leq \epsilon^{-\kappa_{16}/d}$  then  $\mu$  and  $\nu$  are  $\epsilon^{\kappa_{17}/d}$ -close:*

$$|\mu(f) - \nu(f)| \ll_d \epsilon^{\kappa_{17}/d} \mathcal{S}_d(f) \text{ for } f \in C_c^\infty(X).$$

This result is an effective version of the statement “ $\mu(xS) > 0$  implies that  $\mu$  is the  $S$ -invariant measure on the closed orbit  $xS$ .” To prove it, let  $\chi \in C_c(S)$  be a positive compactly supported function on  $S$  with integral one. For a smooth function  $f$  as in (4.3), the values  $\mu(f)$  and  $\mu(f \star \chi)$  are approximately the same by almost invariance. However, repeated applications of the convolution operator  $f \mapsto f \star \chi$  makes  $f$  converge to a constant function, whence the result.

**4.7. The theorem is as easy as ABC.** Say that  $\mu$  is  $[S, \epsilon, d_S]$ -almost invariant if it is  $\epsilon$ -almost invariant under a connected subgroup  $S \supset H$ , w.r.t.  $d_S$ . Propositions D and F, from the previous section, easily convert Propositions A,B,C to the following dichotomy:

**Lemma.** (§16.1) *Suppose that  $\mu$  is  $[S, \epsilon, d_S]$ -almost invariant. There exists constants  $c_2(d_S)$ ,  $\kappa_{18}(d_S)$ ,  $\kappa_{19}(d_S)$ ,  $\kappa_{20}(d_S)$ , and  $d'_S$  so that for any  $\epsilon \leq c_2$  either:*

- $|\mu(f) - \mu_{x_0 S}(f)| \leq c_3(d_S) \epsilon^{\kappa_{19}(d_S)} \mathcal{S}_{d'_S}(f)$ , for some closed orbit  $x_0 S$  of volume  $\leq \epsilon^{-\kappa_{18}(d_S)}$ .

- The measure  $\mu$  is  $[S_*, c_4(d_S)\epsilon^{\kappa_{20}(d_S)}, d'_S]$  almost invariant, where the connected subgroup  $S_* \supset H$  has larger dimension than  $S$ .

Iterating this lemma yields Theorem 1.3.

## 5. BASIC PROPERTIES OF SOBOLEV NORMS.

**5.1. Bounding the  $L^\infty$ -norm.** We extend the notion of Sobolev norm to  $\mathbb{R}^n$ . We define for  $f \in C^\infty(\mathbb{R}^n)$  and the open set  $B_\epsilon = (-\epsilon, \epsilon)^n \subset \mathbb{R}^n$  the Sobolev norm

$$\mathcal{S}_{d, B_\epsilon}(f)^2 = \sum_{|\underline{\alpha}| \leq d} \|f^{(\underline{\alpha})}\|_{L^2(B_\epsilon)}^2,$$

where  $f^{(\underline{\alpha})}$  denotes the  $\underline{\alpha}$ -partial derivative, and the sum is extended over all multi-indexes  $\underline{\alpha}$  of degree  $\leq d$ .

The following establishes (3.9).

**5.1.1. Lemma.** *Let  $n \geq 1$ . There exists a constant  $c(n)$  for which*

$$(5.1) \quad |f(x)| \leq c(n)\epsilon^{-n}\mathcal{S}_{n, B_\epsilon}(f)$$

for all  $\epsilon \leq 1$ ,  $f \in C^\infty(\mathbb{R}^n)$ , and all  $x \in B_\epsilon$ .

Moreover, there exists a constant  $c_1(d)$  and  $\kappa_5$  such that, whenever  $\mathcal{D} \in U(\mathfrak{g})$  has order  $\leq r$ ,

$$(5.2) \quad |\text{ht}(x)^r \mathcal{D}f| \leq c_2(r, \mathcal{D})\mathcal{S}_{d+r}(f)$$

for all  $f \in C_c^\infty(X)$  whenever  $d \geq \kappa_5$ .

*Proof.* The statement in (5.1) follows for  $n = 1$  quite easily: If there is no  $|f(y)|$  smaller than  $\frac{1}{2}|f(x)|$  the statement is clear, otherwise the first derivative will have an integral  $\geq \frac{1}{2}|f(x)|$ . Using Fubini's theorem (5.1) follows by induction on  $n$ .

Let now  $x \in X$  and  $f \in C^\infty(X)$ . We consider the function  $g \mapsto f(xg)$  in a ball of the form  $d(g, 1) \leq \iota_4 \text{ht}(x)^{-\kappa_3}$ , which by (3.7) is an injective image of the corresponding ball in  $G$ . Choosing a coordinate chart for  $G$  around the identity, we transfer the question to  $\mathbb{R}^{\dim G}$  and apply (5.1).  $\square$

**5.2. Trace estimates.** We are going to use the following basic properties of the relative trace. See the paper of Bernstein-Reznikov [2, App. A] for a more careful development of these ideas. The relative trace of two Hermitian forms  $A$  and  $B$  (the latter being positive definite) on a complex vector space  $V$  is defined as

$$\text{Tr}(A|B) = \sum_i \frac{A(e_i)}{B(e_i)}$$

if  $V$  is finite-dimensional and  $e_1, \dots$  is an orthogonal basis of  $V$  with respect to  $B$ .<sup>11</sup> In the infinite dimensional case we define

$$\mathrm{Tr}(A|B) = \sup_{W \subset V} \mathrm{Tr}(A_W|B_W)$$

where the supremum goes over all finite dimensional subspaces  $W$  of  $V$  and  $A_W, B_W$  are the restrictions of  $A$  and  $B$  to  $W$ . Then

$$\mathrm{Tr}(A|B) \leq \mathrm{Tr}(A'|B) \text{ if } A \leq A' \text{ and } \mathrm{Tr}(A|B') \geq \mathrm{Tr}(A|B) \text{ if } B \geq B'$$

for any Hermitian forms  $A, A', B$ , and  $B'$  on  $V$ . Here  $A \leq A'$  means that  $A(v) \leq A'(v)$  for all  $v \in V$ . If  $A \leq B$  then  $\mathrm{Tr}(A|B) = \sum_i A(e_i)$  where  $e_1, \dots$  is an orthonormal basis w.r.t.  $B$ . This was already claimed in §3 and follows from [2, Prop A.2]<sup>12</sup>.

We will be using that notion and these facts also in the case of a general, not necessarily positive definite, Hermitian form  $B$ , but only in the case when  $A$  is zero in the radical of  $B$  (the latter being the subspace where  $B$  is zero). In this case, the trace is defined as the trace on the quotient of  $V$  by the radical.

Although we do not explicitly use it, the following gives a helpful “normalization” of the idea of relative traces. Suppose  $N_1, N_2, \dots, N_r$  are Hermitian norms on  $V$ , so that  $N_i$  is bounded by a multiple of  $N_{i+1}$ ; and the relative trace  $\mathrm{Tr}(N_i|N_{i+1})$  is finite. Then there exists an orthogonal basis  $e_1, \dots, e_k, \dots$  for the completion of  $V$  with respect to  $N_r$ , so that  $N_1$  is also diagonal with respect to this basis, and moreover:

$$(5.3) \quad \sup_j j^{r-1} \frac{N_1(e_j)}{N_r(e_j)} < \infty.$$

**5.3. Proof of (3.10).** For  $x \in X$ , consider the linear form on  $C_c^\infty(X)$  given by:

$$L_x : f \mapsto \mathrm{ht}(x)^r \mathcal{D}f,$$

where  $\mathcal{D}$  is a monomial in the chosen orthonormal basis of  $\mathfrak{g}$  with degree  $\leq r$ . By Lemma 5.1.1,

$$(5.4) \quad \mathrm{Trace}(|L_x|^2 |S_d^2|) < c_2(r, \mathcal{D})^2$$

whenever  $d \geq \kappa_5 + r$ . Integrating (5.4) over  $x \in X$ , and summing over  $\mathcal{D}$ , yields (3.10) (cf. [2, App. A, Prop. 1]).  $\square$

<sup>11</sup>This is well defined since it equals  $\mathrm{Tr}(B^{-1}A)$  if we consider  $A$  and  $B$  as maps to the Hermitian dual  $V^+$  of  $V$ .

<sup>12</sup>In [2, App. A] continuous Hermitian forms on a topological vector space were considered, but with  $A \leq B$  both forms become continuous w.r.t. the norm derived from  $B$ .

## 6. BASIC PROPERTIES PERTAINING TO SPECTRAL GAP.

Our primary aim in this section is to prove the following result, as well as establishing enough background on spectral gap to take advantage of it.

**6.1. Proposition.** *Notation being as in §1.1, let  $H \subset S \subset G$ . Let  $x \in X$ , and  $S_x$  the stabilizer of  $x$  in  $S$ .*

*Then the action of  $S$  on the orthogonal complement to locally constant functions in  $L^2(S_x \backslash S)$  has a spectral gap. Moreover, that spectral gap depends only on  $G, H$ .*

We observe that, in the case when  $H$  has property (T), this result is obvious.

**6.2. The spectral gap and tempered representations.** For general reference on semisimple groups, see [62]. We refer also to [46] where the relation of spectral gap and matrix coefficient decay is discussed and used. Let  $S$  be a (not necessarily connected) semisimple real group with finite center.

**6.2.1. Spectral gap: definition.** We say that a unitary representation  $(\pi, V)$  of  $S$  – not necessarily irreducible – possesses a spectral gap if <sup>13</sup> there is a compactly supported probability measure  $\nu$  on the *connected component* of  $S$ , and  $\delta > 0$ , so that:

$$\|\pi(\nu)v\| < (1 - \delta)\|v\|, \quad (v \in V).$$

It is equivalent to say that the “irreducible constituents of  $\pi$ ,” as a subset of the unitary dual of  $S^0$ , are isolated from the trivial representation in the Fell topology.

**6.2.2. Tempered: definition.** We say that an irreducible, unitary representation  $\pi$  of  $S$  is *tempered* if it is weakly contained in the regular representation  $L^2(S)$ . We say that a unitary representation  $(\pi, V)$  is tempered if it may be disintegrated into tempered representations.

It is equivalent to ask [9, Theorem 1] that there exist a dense subset  $\mathcal{V} \subset V$  so that  $s \mapsto \langle \pi(s)v, w \rangle$  belongs to  $L^{2+\varepsilon}(S)$  for each  $v, w \in \mathcal{V}$  and  $\varepsilon > 0$ .<sup>14</sup> It is also equivalent to ask that the restriction of  $\pi$  to the identity component of  $S$  be tempered; or that the pull-back of  $\pi$  to any finite covering of  $S$  be tempered.

If  $\pi$  is an irreducible representation of  $S$  with compact kernel, then there exists an integer  $m \geq 1$  so that  $\pi^{\otimes m}$  is tempered. This provides a useful

<sup>13</sup>It would be more usual to make this definition without the connected component restriction; the present definition is more convenient for our purposes.

<sup>14</sup>More precisely, that Theorem asserts that, if  $v \in V$  is so that the diagonal matrix coefficient  $\langle \pi(s)v, v \rangle$  belongs to  $L^{2+\varepsilon}$ , then the representation of  $S$  on  $\langle Sv \rangle$  is weakly contained in the regular representation. This means that the diagonal matrix coefficient  $\langle \pi(s)v, v \rangle$  is uniformly approximable, on compacta, by convex combinations of diagonal matrix coefficients associated to the regular representation. Clearly, if this property is valid for a bounded sequence  $v_i$ , it is also valid for any limit of the  $v_i$ s; whence the stated conclusion.

measure how close to tempered a representation is: We say a unitary representation  $\pi$  is *1/m-tempered* if  $\pi^{\otimes m}$  is tempered. (It is equivalent to say that it is  $L^{2m+\varepsilon}$  for all  $\varepsilon > 0$ ; the latter phrasing appears often in the literature.)

**6.2.3. Relations between spectral gap and temperedness.** If  $S$  is almost simple,

(6.1) If  $(\pi, V)$  possesses a spectral gap,  
then  $\pi$  is  $1/m$ -tempered for some  $m \geq 1$ .

This assertion is by now a “folk” result, but is rather remarkable. The property of having a spectral gap is “local” in nature: it depends on the action of group elements near the identity. On the other hand, the property of being  $1/m$ -tempered constrains the action of group elements very far from the identity.

The proof of this assertion may be obtained by combining various results in the literature; see Appendix C. Obviously this is a rather unsatisfactory state of affairs; we don’t know of any entirely conceptual proof.

This ((6.1)) is false if  $S$  is not almost simple. According to our definition, the representation  $1 \otimes L^2(\mathrm{SL}_2(\mathbb{R}))$  of  $\mathrm{SL}_2(\mathbb{R}) \otimes \mathrm{SL}_2(\mathbb{R})$  possesses a spectral gap. However, it is not  $1/m$ -tempered for any  $m \geq 1$ .

**6.3. Some facts on matrix coefficients.** We restrict, in the present section (§6.3), to the case of  $S$  connected.<sup>15</sup>

$S$  admits an Iwasawa decomposition  $S = N.A.K$ ; here  $K$  is a maximal compact subgroup of  $S$  which is the fixed point set of a global Cartan involution  $\Theta : S \rightarrow S$ . Accordingly, there is a projection  $H_A : S \rightarrow A$ .<sup>16</sup>

We define the Harish-Chandra spherical function  $\varphi_0$  via the rule:

$$(6.2) \quad \varphi_0(s) := \int_{k \in K} H_A(ks)^\rho dk$$

where  $\rho : A \rightarrow \mathbb{R}_+$  is the half-sum of the roots (counted with multiplicities) of  $A$  acting on  $N$ , and  $dk$  is the probability Haar measure on  $K$ .

$S$  also admits a Cartan decomposition  $S = KA^+K$ . The function  $\varphi_0$  is bi- $K$ -invariant and belongs to  $L^{2+\varepsilon}(S)$ , for every  $\varepsilon > 0$ . Moreover, in obvious notation (see [30, Proposition 7.15(c)])

$$(6.3) \quad |\varphi_0(ka_+k')| \ll_\varepsilon \rho(a_+)^{-1+\varepsilon},$$

for any  $\varepsilon > 0$ ; the implicit constant here depends on the isomorphism class of  $S$  only.

<sup>15</sup>This is mainly because of the lack of suitable references in general.

<sup>16</sup>The notions we are about to define depend on the choice of  $K$ , although this dependency is not very important. Moreover, we use the results of this section for  $S$  an intermediate group between  $H$  and  $G$ ; for such we have fixed a choice of maximal compact after Lemma 3.4.1.

Moreover, for any one-parameter unipotent subgroup  $u : \mathbb{R} \rightarrow S$  we have the bound <sup>17</sup>

$$(6.4) \quad |\varphi_0(u(t))| \ll_\varepsilon (1 + |t|)^{-1+\varepsilon}.$$

It will be convenient to make a slight generalization of this definition to handle some slight complications arising from groups with multiple simple factors. If  $S$  is a direct product  $S_1 \times \dots \times S_k$  of almost simple groups, we set:

$$(6.5) \quad \varphi_0^{\text{wk}}(s_1, \dots, s_k) = \max_{1 \leq i \leq k} \varphi_{0, S_i}(s_i).$$

Even if  $S$  fails to be a direct product, i.e. is isogenous to a direct product, there exists an isogeny  $\prod_{i=1}^I S_i \rightarrow S$ ; the function  $\varphi_0^{\text{wk}}$  defined on  $\prod_{i=1}^I S_i$  by (6.5) is bi-invariant by a maximal compact of  $\prod_{i=1}^I S_i$ , and in particular by the kernel of the isogeny. It thereby descends to  $S$ .

**6.3.1. Lemma.** *Let  $F \subset S$  be bi- $K$ -invariant, i.e.  $K.F.K = F$ . Then*

$$(6.6) \quad \text{vol}(F)^{-1} \int_F \varphi_0(s) ds \ll \text{vol}(F)^{-1/3},$$

$$\text{vol}(F)^{-2} \int_{s, s' \in F} \varphi_0(ss'^{-1})^{1/p} \ll \text{vol}(F)^{-\frac{2}{3p}}.$$

*The implicit constants in these estimates depend only the isomorphism class of  $S$  and the choice of Haar measure.*

*Proof.* The first assertion follows by the duality between  $L^3$  and  $L^{3/2}$ .

Next, let  $p \geq 1$ , and consider  $\int_F \varphi_0(ss'^{-1})^{1/p} ds ds'$ . Let  $q$  satisfy  $1/p + 1/q = 1$ ; then, by duality between  $L^p$  and  $L^q$ ,

$$\text{vol}(F)^{-2} \int_{s, s' \in F} \varphi_0(ss'^{-1})^{1/p} \leq \left( \int_{s, s' \in F} \varphi_0(ss'^{-1}) \right)^{1/p} \text{vol}(F)^{-2/p}$$

Noting the identity  $\int_{k \in K} \varphi_0(s_1 k s_2) = \varphi_0(s_1) \varphi_0(s_2)$ , cf. [30, (7.45)], we simplify this to (6.6).  $\square$

**6.3.2. Bounds for matrix coefficients of tempered and  $1/m$ -tempered representations.** If  $(\pi, V)$  is a tempered representation of  $S$ , it is known [9] that:

$$(6.7) \quad \langle s.v, w \rangle \leq \varphi_0(s) \cdot \|v\| \|w\| (\dim Kv)^{1/2} \cdot (\dim Kw)^{1/2} \quad (v, w \in V)$$

Fix a basis for the Lie algebra  $\mathfrak{s}$  of  $S$ . We may define a system of Sobolev norms  $\mathcal{S}_d^V$  on the smooth subspace<sup>18</sup> of any unitary representation  $V$  via<sup>19</sup>

$$\mathcal{S}_d^V(f) := \sup_{\mathcal{D}} \|\mathcal{D}f\|$$

<sup>17</sup>As can be deduced, e.g., by embedding  $u$  in an  $\text{SL}_2(\mathbb{R})$ .

<sup>18</sup>Here  $v \in V$  is smooth if the associated orbit map  $G \xrightarrow{g \mapsto g.v} V$  is smooth,

<sup>19</sup>It should be noted that, when specialized to the case of  $V = L_\mu^2(\Gamma \backslash G)$ , this gives norms that do not coincide with the family of norms defined in (3.8). However, the latter majorizes the former if we allow different indices and use (3.9) for as many derivatives along  $\mathfrak{s}$  as needed.

where the supremum is taken over  $\mathcal{D} \in U(\mathfrak{g})$ , the universal enveloping algebra of  $\mathfrak{s}$ , which are monomials in the chosen basis of degree  $\leq d$ . Then

$$(6.8) \quad \langle s.v, w \rangle \leq c_k \varphi_0(s) \cdot \mathcal{S}_k^V(v) \mathcal{S}_k^V(w) \quad (v, w \in V)$$

for  $s \in S$ , any integer  $k > (\dim K)/2$  and a constant  $c_k \geq 1$ .

Suppose, now, that  $(\pi, V)$  is  $1/m$ -tempered; in that case we have:

$$(6.9) \quad \langle s.v, w \rangle \leq c'_k \varphi_0(s)^{1/m} \mathcal{S}_k^V(v) \mathcal{S}_k^V(w) \quad (v, w \in V)$$

for  $s \in S$ .

**6.3.3. Bounds for matrix coefficients in presence of a spectral gap.** As we have discussed (end of §6.2.3) it is possible for a representation to possess a spectral gap, but not to be  $1/m$ -tempered for any  $m \geq 1$ . In order to quantify the decay of matrix coefficients in such cases, we use the function  $\varphi_0^{\text{wk}}$  of (6.5).

If  $(\pi, V)$  possesses a spectral gap, there exists  $\rho > 0$ , depending only on this gap, such that the following majorization holds:

$$(6.10) \quad \langle s.v, w \rangle \leq c_k (\varphi_0^{\text{wk}}(s))^\rho \cdot \mathcal{S}_k^V(v) \mathcal{S}_k^V(w) \quad s \in S, (v, w \in V),$$

for  $k > (\dim K)/2$ .

By unitary decomposition it suffices to verify (6.10) for every irreducible “constituent” of  $V$ , i.e. every irreducible unitary representation that weakly occurs within  $V$ .

Take an isogeny  $\prod_{i=1}^I S_i \rightarrow S$  with each  $S_i$  connected almost simple. Any irreducible unitary representation of  $\prod_{i=1}^I S_i$  factors as a tensor product  $\otimes_{i=1}^I \sigma_i$ , where each  $\sigma_i$  is irreducible.

The assumption of spectral gap implies (cf. (6.1)) that there exists an integer  $m$  so that every irreducible “constituent”, upon pullback to  $\prod_{i=1}^I S_i$ , is of the form  $\sigma = \otimes_{i=1}^I \sigma_i$ , where there exists  $1 \leq i \leq I$  so that  $\sigma_i^{\otimes m}$  is tempered as an  $S_i$ -representation. Assume  $i = 1$ , the argument being similar in general.

For  $v, w$  in the space of  $\sigma$ , put  $w' := (1, s_2, \dots, s_I)^{-1}w$ . Then:

$$(6.11) \quad \begin{aligned} \langle (s_1, \dots, s_I)v, w \rangle &= \langle (s_1, 1, \dots, 1)v, (1, s_2, \dots, s_I)^{-1}w \rangle \\ &\ll \varphi_0(s_1)^{1/m} S_k(v) S_k(w') \leq \varphi_0^{\text{wk}}((s_1, \dots, s_k))^{1/m} S_k(v) S_k(w). \end{aligned}$$

The very last step follows, because the Sobolev norms here are taken on  $\sigma$  considered as an  $S_1$ -representation, and in particular commute with the action of  $S_2 \times \dots \times S_I$ . Our assertion follows.

**6.4. Some estimates of volumes and matrix coefficients.** In the present section,  $H, G$  are as in §1.1.

Let  $H \subset S \subset G$ , with  $S$  connected. We claim that for  $T \geq 2$ , and for any  $\rho > 0$ ,

$$(6.12) \quad \text{vol } B_S(T) \sim v_S T^{A_S} (\log T)^{\ell_S},$$

$$\frac{1}{\text{vol } B_S(T)^2} \int_{g, g' \in B_S(T)} \varphi_0^{\text{wk}}(gg'^{-1})^\rho \ll T^{-\zeta_S \rho}$$

for suitable  $v_S, A_S, \zeta_S > 0$ , and  $\ell_S \geq 0$  depending on  $S$ . Here  $f(t) \sim g(t)$  if  $\frac{f(t)}{g(t)} \rightarrow 1$  as  $t \rightarrow \infty$ .

The estimate on the volume follows from [25, Thm. 2.7]. (Actually, we do not need this level of precision, but it is convenient to have an asymptotic.)

Indeed, notations as in (3.5), it suffices to compute the measure of the inverse image of  $B_S(T)$  under  $\Phi$ , i.e. the measure of

$$(6.13) \quad \ker(\Phi) \cdot \prod_i \{s \in S_i : \|\Phi(s_i)\| \leq N^{-1} T^{1/I}\}.$$

We can regard this as a certain union of sets, parameterized by the finite kernel  $\ker(\Phi)$ . The results stated in [25] implies an asymptotic for each intersection of these sets, whence also an asymptotic for their union.<sup>20</sup>

Now, let us indicate the proof of the second assertion of (6.12). Take the isogeny  $\Phi : \prod_{i=1}^I S_i \rightarrow S$  from (3.5). Because  $\max(a^\rho, b^\rho) \ll a^\rho + b^\rho$ , and the Haar measure on  $S$  pulls back to the Haar measure on  $\prod_i S_i$ , it is enough to check, for each  $1 \leq i \leq I$ , the average value of  $\varphi_0(s_i s_i'^{-1})^\rho$  over (6.13) is bounded by a negative power of  $T$ . The set (6.13) is contained in a set of the form

$$(6.14) \quad \prod_i \{s \in S_i : \|\Phi(s_i)\| \leq R\}.$$

where  $R \asymp T^{1/I}$ ; moreover, the sets (6.13) and (6.14) have comparable volumes, by what we have just discussed.

We conclude that it is enough to check that, for each  $1 \leq i \leq I$ , the average of  $\varphi_0(ss'^{-1})^\rho$  over the set  $\{(s, s') \in S_i, \|\Phi(s)\|, \|\Phi(s')\| \leq R\}$ , is bounded above by a negative power of  $R$ .

That assertion, however, is an almost-immediate consequence of (6.6); we need only enlarge  $\{s \in S_i : \|\Phi(s)\| \leq R\}$  so that it is bi-invariant under a maximal compact of  $S_i$ . This increases the volume by at most a constant factor, by property (1) of (3.5) and the above volume estimate.

### 6.5. The representation of $\mathbf{S}(\mathbb{R})$ on an algebraic homogeneous space.

Having established basic definitions concerning tempered and  $\frac{1}{m}$ -tempered representations, we now show that certain naturally occurring examples, viz., representations of the real points of an algebraic group, on the real points of an (algebraic) homogeneous spaces, have such properties.

<sup>20</sup>The reference [25] assumes that  $S$  connected. It also proves “Theorem 2.7” only in a special case, deferring the general case to another paper. We note however that Benoist and Oh [1] have given a lovely conceptual argument for such volume asymptotics.

We say that, if  $V$  is a vector space over a field  $k$ , and  $\mathbf{G} \subset \mathrm{GL}(V)$  an algebraic group, then  $v \in V$  is *stable* if it is not contained entirely in the non-negative weight spaces for any  $k$ -split torus  $\mathbf{G}_m \subset \mathbf{G}$ .

For example, if  $\mathbf{G} = \mathrm{SL}_2$ , the irreducible (algebraic) representations of  $\mathbf{G}$  are precisely the representations on the space of degree  $m$  bivariate homogeneous polynomials, for  $m \geq 1$ . According to the definition above, a degree  $m$  polynomial is then stable exactly when there is no root in  $\mathbb{P}^1(k)$  of multiplicity  $\geq m/2$ .

We will use this notion when  $k = \mathbb{R}$ . In this case, the set of stable points is topologically open.

**6.5.1. Lemma.** *Suppose  $S$  acts properly on a real manifold  $M$ . Then, if  $\nu$  is an  $S$ -invariant Radon measure on  $M$ , then  $L^2(M, \nu)$  is tempered as an  $S$ -representation.*

Recall *properness* means that the map  $S \times M \rightarrow M \times M$ , defined via  $(s, m) \mapsto (sm, m)$  is proper.

*Proof.* Indeed, it suffices to verify that matrix coefficients  $s \mapsto \langle sf_1, f_2 \rangle$ , when  $f_1, f_2 \in C_c(M)$ , are in  $L^{2+\varepsilon}(S)$ . By properness, they are compactly supported.  $\square$

**6.5.2. Lemma.** *Suppose we are given a finite-dimensional algebraic representation  $(\rho, V)$  of  $\mathrm{SL}_2$ . Let  $M \subset V_{\mathbb{R}}$  be a  $\mathrm{SL}_2(\mathbb{R})$ -stable submanifold so that every  $m \in M$  is stable.*

*Then the action of  $\mathrm{SL}_2(\mathbb{R})$  on  $M$  is proper.*

*Proof.* Let  $A = \{a_t\}$  be a diagonal torus within  $\mathrm{SL}_2(\mathbb{R})$ . It suffices to verify that the action of  $A$  on  $M$  is proper, in view of the Cartan decomposition.

Split  $V = \oplus V_n$ , where  $\{a_t\}$  acts on  $V_n$  by the character  $e^{nt}$ .

Choose compacta  $K_1, K_2 \subset M$ . Suppose, to the contrary, that there exists a sequence  $u_i \in K_1, v_i \in K_2$  and an unbounded sequence  $a_{t_i} \in A$  so that  $v_i = a_{t_i} u_i$ . Let  $u$  be a limit of the  $u_i$  and  $v$  a limit of the  $v_i$ . Then neither  $u$  nor  $v$  are stable, a contradiction.  $\square$

**6.5.3. Lemma.** *Suppose we are given a finite-dimensional algebraic representation  $(\rho, V)$  of  $\mathrm{SL}_2$ . Let  $\nu$  be an  $\mathrm{SL}_2(\mathbb{R})$ -invariant Radon measure on  $V_{\mathbb{R}}$  so that  $\nu$ -almost all vectors  $v \in V_{\mathbb{R}}$  are not fixed by  $\mathrm{SL}_2(\mathbb{R})$ . Then almost all vectors in  $V_{\mathbb{R}} \oplus V_{\mathbb{R}}$ , w.r.t.  $\nu \times \nu$ , are stable.*

*Proof.* If  $v_1, v_2 \in V_{\mathbb{R}}$  are not fixed by  $\mathrm{SL}_2(\mathbb{R})$ , then  $v_1 \oplus gv_2$  is stable in  $V \oplus V$  for almost all  $g \in \mathrm{SL}_2(\mathbb{R})$ . This, together with Fubini's theorem, implies the stated claim.  $\square$

**6.5.4. Lemma.** *Let  $\mathbf{S}$  be a semisimple  $\mathbb{R}$ -algebraic group without anisotropic factors, and let  $\mathbf{L} \subset \mathbf{S}$  be an algebraic subgroup of strictly lower dimension. Let  $S$  resp.  $L$  be the real points of  $\mathbf{S}$  resp.  $\mathbf{L}$ . Then the right action of  $S$  on  $L^2(L \backslash S)$  has a spectral gap.*

Although this case does not arise in our application, one should interpret  $L^2(L \backslash S)$  as the “unitary induction from  $L$  to  $S$  of the trivial representation,” in the case where there fails to be an  $S$ -invariant measure on  $L \backslash S$ .

*Proof.* It is easy to see that one may assume that  $\mathbf{L}$  has no characters over  $\mathbb{R}$ . Otherwise, replace  $\mathbf{L}$  by the intersection  $\mathbf{L}'$  of the kernels of all such characters. Then  $L^2(L' \backslash S)$  weakly contains  $L^2(L \backslash S)$ , and it suffices to prove the theorem replacing  $\mathbf{L}$  by  $\mathbf{L}'$ . Similarly, we may assume that  $\mathbf{L}$  is connected.

Fix a morphism from  $\mathrm{SL}_2$  to  $\mathbf{S}$  that projects nontrivially to any almost-simple component of  $\mathbf{S}$ . Then the measure of points in  $L \backslash S$  fixed by  $\mathrm{SL}_2(\mathbb{R})$  is zero.

There exists a representation  $(\rho, V)$  of  $\mathbf{S}$  and a rational vector  $v_L \in V$  so that the stabilizer of the line  $\mathbb{R} \cdot v_L$  is precisely  $\mathbf{L}$ . By assumption,  $\mathbf{L}$  is connected and has no  $\mathbb{R}$ -characters, so  $\mathbf{L}$  is the stabilizer of  $v_L$ .

Let  $\mathbf{Y}$  be the orbit of  $v_L$  under  $\mathbf{S}$ ; it is an algebraic subvariety of  $V$ . Let  $Y = \mathbf{Y}(\mathbb{R})$ . Then  $L \backslash S$  is identified with an open subset of  $Y$ . Let  $\nu$  be the measure on  $Y$  corresponding to an  $S$ -invariant measure on  $L \backslash S$ , which exists since  $L$  is unimodular.

Consider  $V \oplus V$  as an  $\mathrm{SL}_2$ -representation. Consider  $Y \times Y \subset V_{\mathbb{R}} \times V_{\mathbb{R}}$ . The set of stable points  $M \subset Y \times Y$  is an open subset of full measure (Lemma 6.5.3). So  $L^2(Y \times Y, \nu \times \nu) = L^2(M, \nu \times \nu)$ .

We apply Lemma 6.5.1–6.5.3 to conclude that the  $\mathrm{SL}_2(\mathbb{R})$ -action on the tensor product  $L^2(L \backslash S)^{\otimes 2}$  is tempered. Therefore, the  $S$ -action on  $L^2(L \backslash S)$  has a spectral gap.  $\square$

## 6.6. Property $\tau$ and its corollaries.

**6.7. Proposition.** *Let  $\mathbf{S}$  be an absolutely almost simple, simply connected group over a number field  $F$ . Let  $v$  be a place of  $F$ . There is an integer  $q_{\mathbf{S}(F_v)}$ , which depends only on the isomorphism class of  $\mathbf{S}(F_v)$ , so that the representation  $\mathbf{S}(F_v)$  on  $L_0^2(\mathbf{S}(F) \backslash \mathbf{S}(\mathbb{A}_F))$  is  $\frac{1}{q_{\mathbf{S}(F_v)} - 1}$ -tempered.*

In fact, the integer  $q$  may be taken to depend only on  $\dim(\mathbf{S})$ , but we do not need this. We say “ $\frac{1}{q-1}$ -tempered”, rather than “ $\frac{1}{q}$ -tempered”, to absorb some annoying factors of  $\varepsilon$  at later stages in our proof (examples of such factors can be seen in (6.4)).

*Explicit* (and rather “good”) values for  $q$  could be derived from the work of H. Oh and Gorodnik-Maucourant-Oh: [47] and [24, Cor 3.26]; these results give explicit (and “good”) rates of decay for matrix coefficients. Indeed, in the nonarchimedean case when the rank is  $\geq 2$ , [47] is apparently the only place in the literature where the existence of  $q$  is established.

*Proof.* We confine ourselves to the case of  $v$  archimedean, which is the only case we use in the present paper (cf. remarks above).

This follows from the solution to property  $\tau$  which was completed by L. Clozel [6]; this solution uses a variety of ingredients: the trace formula, prior

ideas of Burger and Sarnak [5], the ideas of Kazhdan on property  $T$ , and work of A. Selberg for groups of type  $A_1$ .

However, it does not follow from the main statement of [6], but rather from the proof. Indeed, the statements of [6] *a priori* depend on the  $F$ -form  $\mathbf{G}$ , and not just the group  $G$ ; however, the proofs give stronger statements.

If the real rank  $\mathbf{S}(F_v)$  exceeds 1, the assertion is proved explicitly in [8, Theorem 2.4]. If the real rank of  $S := \mathbf{S}(F_v)$  equals 1, it is proven in [6] that there exists a homomorphism  $H \rightarrow S$ , where  $H$  is a real Lie group locally isomorphic to  $\mathrm{SU}(n, 1)$ , some  $n \geq 1$ , with the property that the pull-back of  $L_0^2(\mathbf{S}(F) \backslash \mathbf{S}(\mathbb{A}_F))$  is isolated from the identity as an  $H$ -representation. Moreover, this notion of “isolated” is absolute, i.e. independent of  $\mathbf{S}$ .

However, there exist only finitely many conjugacy classes of homomorphisms (of real Lie groups) from  $\mathrm{SU}(n, 1)$  to  $S$ ; see Lemma A.1.1. It follows from that statement (cf. (6.1)) that the  $S$ -action on  $L_0^2$  has a spectral gap depending only the isomorphism class of the real Lie group  $S$ , and therefore there exists an integer  $q_S$  as asserted.  $\square$

**6.7.1. Lemma.** *Notation being as in §1.1, let  $H \subset S \subset G$ ; let  $xS$  be a closed connected orbit of  $S$  on  $\Gamma \backslash G$ . Let  $\nu$  be the  $S$ -invariant probability measure on  $xS$ .*

*There exists  $p_G \geq 1$ , which can be taken to depend only on  $\dim G$ , so that  $L_0^2(\nu)$  is  $\frac{1}{p_G-1}$ -tempered as an  $S$ -representation.*

Note that this means that  $L_0^2(\nu)$  is uniformly isolated (as one varies the closed orbit  $xS$ ) from the identity as an  $S$ -representation or as an  $H$ -representation.

*Proof.* Replace  $S$  by  $xS^0x^{-1}$  and  $\Gamma xS$  by  $\Gamma xS^0x^{-1}$ . It is easy to see that it suffices to prove the statement in this setting.

By the Borel-Wang density theorem, [37, Chapter II, Corollary 4.4] there exists a semisimple connected  $\mathbb{Q}$ -group  $\mathbf{S}$  so that  $S$  is a finite index subgroup of  $\mathbf{S}(\mathbb{R})$ . (Take the Zariski closure of  $\Gamma \cap S$ ).

Set  $\Gamma_S = \Gamma \cap \mathbf{S}(\mathbb{R})$ . It is a congruence subgroup of  $\mathbf{S}(\mathbb{Q})$ . It suffices to verify that the representation of  $\mathbf{S}(\mathbb{R})$  on  $L_0^2(\Gamma_S \backslash \mathbf{S}(\mathbb{R}))^{\otimes m}$  is tempered. (Here  $L_0^2$  denotes orthogonal complement of *locally* constant functions.)

$\Gamma_S$  being a congruence subgroup, it is enough to check that the representation of  $\mathbf{S}(\mathbb{R})$  on  $L_0^2(\mathbf{S}(\mathbb{Q}) \backslash \mathbf{S}(\mathbb{A}))^{\otimes m}$  is tempered. In this context, we understand  $L_0^2$  as meaning the orthogonal complement of  $\mathbf{S}(\mathbb{R})^0$ -invariant functions.

There exists number fields  $K_i$ , and absolutely almost simple, simply connected groups  $\mathbf{S}_i$  over  $K_i$ , together with an isogeny  $\varphi : \prod_{i=1}^k \mathrm{Res}_{K_i/\mathbb{Q}} \mathbf{S}_i \rightarrow \mathbf{S}$ . It is then enough to verify that the representation of  $\mathbf{S}_i(K_i \otimes \mathbb{R})$  on  $L_0^2(\mathbf{S}_i(K_i) \backslash \mathbf{S}_i(\mathbb{A}_{K_i}))^{\otimes m}$  is tempered. Indeed, this guarantees that the representation of the identity component of  $\prod_i \mathbf{S}_i(K_i \otimes \mathbb{R})$  on  $L_0^2(\mathbf{S}(\mathbb{Q}) \backslash \mathbf{S}(\mathbb{A}))^{\otimes m}$

is tempered. But the former group is a finite covering of the identity component of  $\mathbf{S}(\mathbb{R})$ . We apply Proposition 6.7 to conclude, taking for  $p_G$  the maximum of all the integers  $q_{\mathbf{S}(\mathbb{R})}$  associated to possible  $\mathbf{S}$ s.  $\square$

**6.8. Proof of Proposition 6.1.** Recall that for a subgroup  $S$  of  $G$  we denote the normalizer of the Lie algebra of  $S$  by  $\tilde{S}$ .

It suffices to show that the action of  $S$  on the orthogonal complement to constant functions in  $L^2(\tilde{S}_x \backslash \tilde{S})$  has a spectral gap; or, that the action of  $\tilde{S}$  on the same space has a spectral gap. Set  $\mathbf{L}$  to be the Zariski closure of  $\tilde{S}_x$  inside  $G$ ; then let  $L = \mathbf{L}(\mathbb{R}) \subset \tilde{S}$ . If  $L = \tilde{S}$ , the  $\tilde{S}$ -orbit of  $\Gamma g \in \Gamma \backslash G$  is closed, and we apply Proposition 6.7.1. Otherwise  $L$  satisfies  $\dim L < \dim S$ .<sup>21</sup>

The representation  $L^2(\tilde{S}_x \backslash \tilde{S})$  may be regarded as the induction from  $L$  to  $\tilde{S}$ , of  $L^2(\tilde{S}_x \backslash L)$ . Now, if  $G_1 \subset G_2$  are locally compact groups,  $V$  a unitary representation of  $G_1$ , and  $\nu$  a probability measure on  $G_2$ , then the operator norm of convolution with  $\nu$  on  $\text{Ind}_{G_1}^{G_2} V$  is bounded by the corresponding norm on  $L^2(G_1 \backslash G_2)$ . In particular, if  $L^2(G_1 \backslash G_2)$  has a spectral gap, so also does  $\text{Ind}_{G_1}^{G_2} V$ . We invoke Lemma 6.5.4 to conclude.  $\square$

## 7. EFFECTIVE GENERATION OF LIE ALGEBRAS.

Given a subset  $T = \{t_1, \dots\}$  of a Lie algebra  $\mathfrak{g}$ , recall that we denote by  $T^{(k)}$  the set of all possible iterated Lie brackets of the  $t_i$ s of depth  $\leq k$ . In this section, we shall prove the following (“Proposition E”):

**7.1. Proposition.** *There exists an integer  $k$  and some  $c > 0$  depending only on  $\mathfrak{g}$  with the following property. For any orthonormal subset  $T = \{t_1, \dots\}$  of  $\mathfrak{g}$ , and  $0 < \delta < 1$  there exists a subalgebra  $\mathfrak{w} \subset \mathfrak{g}$  with orthonormal basis  $w_1, w_2, \dots$  satisfying:*

- (1) *For each  $w_i$ , there exists a linear combination  $w'_i = \sum_{t \in T^{(k)}} c_t t$ , with  $c_t \in \mathbb{R}$  satisfying  $|c_t| \leq \delta^{-k}$ , so that  $\|w_i - w'_i\| \ll \delta^c$ .*
- (2) *Each  $t \in T$  is within  $\delta$  of  $\mathfrak{w}$  (i.e.  $\min_{w \in \mathfrak{w}} \|w - t\| \leq \delta$ ).*

*If the linear span  $\langle T \rangle$  of  $T$  contains a subalgebra  $\mathfrak{h}$ , then there exists  $k$  and  $c$  depending on the pair  $(\mathfrak{h}, \mathfrak{g})$  so that (1) and (2) hold and, in addition,  $\mathfrak{h} \subset \mathfrak{w}$ .*

Let us discuss the motivation for this. Given elements  $e_1, \dots, e_r \in \mathfrak{g}$ , one can speak of the “subalgebra  $\mathfrak{w}$  spanned by  $\mathfrak{g}$ ”, but this notion is not robust under perturbation: this subalgebra might drop drastically under a small modification of the  $e_i$ . This is related to the fact that the  $e_i$  might span  $\mathfrak{w}$  very *inefficiently*. In explicit terms, there may exist  $w \in \mathfrak{w}$  of small norm, which cannot be written as a linear combination of Lie brackets of the  $e_i$ s with small coefficients. Proposition 7.1 constructs an effective replacement: We construct a subalgebra  $\mathfrak{w}$  which “almost” contains  $T$  and

<sup>21</sup> Note that the group  $\mathbf{L}$  conjugated by an element of  $G$  representing  $x$  is defined over  $\mathbb{Q}$  and is  $\mathbb{Q}$ -anisotropic since its integer points are Zariski-dense. In particular, its real points are unimodular.

so that elements of  $T^{(k)}$ , for suitable small<sup>22</sup>  $k$ , “almost span  $\mathfrak{w}$  in an efficient fashion.”

The proof, given in §7.4 will require some setup on Euclidean spaces (§7.2) and a recollection of the Łojasiewicz inequality (§7.3).

**7.2.** Suppose  $V$  is a Euclidean vector space. We shall say that two subspaces  $U_1, U_2 \subset V$  are  $\epsilon$ -close if there are orthonormal bases  $u_1, \dots, u_r$  and  $v_1, \dots, v_r$  for  $U_1, U_2$  so that  $\|u_i - v_i\| \leq \epsilon$ . We say that  $U_1$  is  $\epsilon$ -almost contained in  $U_2$  if there exists  $U'_2 \subset U_2$  so that  $U \stackrel{\epsilon}{\sim} U'_2$ .

Suppose that there exists an orthonormal basis  $u_1, \dots, u_r$  of  $U_1$  so that  $\text{dist}(u_i, U_2) \leq \epsilon$ . Then there exists a constant  $C_V$ , depending only on  $\dim V$ , so that  $U_1$  is  $C_V \epsilon$ -almost contained in  $U_2$ .

Suppose  $V, W$  are two Euclidean vector spaces and  $f : V \rightarrow W$  a linear map between them. We recall the singular value decomposition: let  $w_1, \dots, w_n$  be an orthonormal basis of  $W$  consisting of eigenvectors of  $f f^t$  with decreasing eigenvalues where  $f^t : W \rightarrow V$  denotes the transpose map. Then the eigenvalues  $(f f^t w_i, w_i) = \|f^t w_i\|^2$  are nonnegative, and we define  $v_i = \frac{1}{\|f^t w_i\|} f^t w_i$  whenever this is defined, say for  $i \leq k$ , and extend it to an orthonormal basis of  $V$ . With this choice we have  $(f v_i, w_i) = \frac{1}{\|f^t w_i\|} (f f^t w_i, w_i) = \|f^t w_i\| = \|f v_i\| = \sigma_i$  for  $i \leq k$  and  $f v_i = 0$  otherwise.

Therefore, there are orthonormal bases  $v_1, v_2, \dots, v_m$  for  $V$  respectively  $w_1, w_2, \dots, w_n$  for  $W$  so that  $f(v_i) = \sigma_i w_i$  for some  $\sigma_i \geq 0$ ; here, by definition,  $w_i = 0$  for  $i > n$ . Let  $W[\delta]$  be the space spanned by those  $w_i$ ,  $i \leq k$ , for which  $|\sigma_i| \geq \delta$ .

Then, for any  $v \in V$ :

$$(7.1) \quad \text{the element } f(v) \text{ is within } \delta \|v\| \text{ of } W[\delta].$$

To see this, let  $v = \sum_i t_i v_i$  and write  $f(v) = \sum_{i: |\sigma_i| \geq \delta} \sigma_i t_i w_i + \sum_{i: |\sigma_i| < \delta} \sigma_i t_i w_i$ . Then the second sum has norm  $\leq \delta \|v\|$ , as claimed.

We will be using this construction for various choices of the function  $f$  and the domain  $V$  to obtain various subspaces of  $W = \mathfrak{g}$ , and the following remark will help to compare their dimension. Let  $V_1 \subset V$  be a subspace, then

$$(7.2) \quad \|f(v)\| \geq \delta \|v\| \text{ for all } v \in V_1 \text{ implies } \dim W[\delta] \geq \dim V_1.$$

Suppose, to the contrary, that  $\dim W[\delta] < \dim V_1$ . Then there would exist  $v \in V_1$  perpendicular to all the  $v_i$  with  $|\sigma_i| \geq \delta$ . This gives a contradiction to (7.1) since  $f(v)$  is perpendicular to  $W[\delta]$  and has length  $\|f(v)\| \geq \delta \|v\|$  by assumption.

**7.3. Łojasiewicz inequality.** The *Łojasiewicz inequality* states that, for  $U \subset \mathbb{R}^n$  open,  $K \subset U$  a compact set, and  $f : U \rightarrow \mathbb{R}$  a real analytic function with zero-set  $Z_f$ , there exist constants  $c_1, c_2 > 0$  so that:

$$(7.3) \quad |f(x)| \geq c_1 \text{dist}(x, Z_f)^{c_2}, x \in K$$

<sup>22</sup>Here the word ‘small’ refers to the fact that it only depends on  $\mathfrak{g}$ .

Here  $\text{dist}$  refers to the Euclidean distance on  $\mathbb{R}^n$ . For this, see [35, Theorem 4.1].

**7.4. Proof of Prop. 7.1.** We define the map  $f_m : \mathbb{R}^{T(m)} \rightarrow \mathfrak{g}$  by sending  $(v_t)_{t \in T(m)}$  to  $\sum_{t \in T(m)} v_t t$ . Considering  $\mathbb{R}^{T(m)}$  as an Euclidean space in the usual way, we obtain the subspace  $W_m[\delta] \subset W = \mathfrak{g}$  by the above construction.

From the definition it follows that the spaces  $W_m[\delta]$  increase when  $\delta$  decreases and  $m$  is held fixed. Moreover, their *dimensions* increase when  $m$  increases and  $\delta$  is held fixed, which follows from (7.2). Finally, as follows from (7.1): if  $m_1 \leq m_2$ , then there is an orthonormal basis for  $W_{m_1}[\delta]$ , all of whose elements are within  $\epsilon \delta^{-1}$  of  $W_{m_2}[\epsilon]$ , i.e.  $W_{m_1}[\delta]$  is  $\ll \epsilon \delta^{-1}$ -almost contained in  $W_{m_2}[\epsilon]$ . In particular, every  $t \in T$  is within  $\epsilon$  of  $W_m[\epsilon]$  for each  $m \geq 1$ .

We claim there is a constant  $\kappa$  depending only on  $\dim(G)$ , so that, for any  $\delta \in (0, 1)$ , we may find  $\delta_1 \in [\delta^\kappa, \delta]$  and  $m \leq \kappa$  so that

$$\dim W_m[\delta_1] = \dim W_{2m}[\delta_1^3]$$

In fact, this follows from the above remarks: set initially  $\delta_1 = \delta$  and  $m = 1$ , and consider the dimensions  $\dim W_m[\delta_1] \leq \dim W_{2m}[\delta_1^3]$ . If equality holds, we are done. Otherwise set  $\delta'_1 = \delta_1^3$  and  $m' = 2m$ . This way we have increased the dimension of  $W_{m'}[\delta'_1]$ . Since this can only happen  $\dim(G)$  often if we repeat the process, the claim follows.

Applying (7.1) once more we get that  $W := W_m[\delta_1]$  has an orthonormal basis  $w_1, \dots, w_r$  so that  $[w_i, w_j]$  is at distance  $\ll \epsilon \delta_1^{-2}$  from  $W_{2m}[\epsilon]$  for any  $\epsilon > 0$ . In particular,  $[w_i, w_j]$  is at distance  $\ll \delta$  from  $W$ .

It follows that there exists a Lie subalgebra  $\mathfrak{w} \subset \mathfrak{g}$  satisfying  $\mathfrak{w} \stackrel{\delta^c}{\approx} W$ , where  $c$  depends only on  $\mathfrak{g}$ . To see this set  $r = \dim W$ . Consider the open subset  $U \subset \mathfrak{g}^r$  consisting of vectors  $(X_1, \dots, X_r)$  which are linearly independent. The inner product on  $\mathfrak{g}$  induces one on all exterior powers. The real-analytic function

$$(7.4) \quad F := \sum_{i,j} \|X_1 \wedge \dots \wedge X_r \wedge [X_i, X_j]\|^2$$

vanishes precisely when the span  $\langle X_1, \dots, X_r \rangle$  is a subalgebra. Clearly our assumptions imply that  $|F(w_1, \dots, w_r)| \ll \delta^2$ ; applying (7.3) to a suitable compact inside  $U$  yields the result.

For the final assertion, we apply the Lojasiewicz inequality to a slightly different variety. In that case the subspace  $W$  has the following additional property:  $\mathfrak{h}$  is  $\ll \delta$ -almost contained in  $W$ . This means that  $W$  almost belongs to the closed subvariety of the Grassmannian defined by subspaces  $\mathfrak{w} \subset \mathfrak{g}$  which are Lie subalgebras and which contain  $\mathfrak{h}$ . Invoking (7.3) for a suitably modified version of (7.4) concludes the proof.  $\square$

## 8. ALMOST INVARIANCE OF MEASURES.

We begin with some reminders to the reader about our notations. These points were discussed in §3, but they are particularly pertinent here. Firstly, we always assume that any Sobolev norm  $\mathcal{S}_d$  we use satisfies  $d \geq \kappa_5 + 1$ , i.e., involves enough derivatives that (3.13) is valid. Secondly, if  $X \in \mathfrak{g}$ , the notions of “almost invariant under  $\exp(X)$ ” and “almost invariant under  $X$ ” do not coincide; the latter is stronger, as it entails almost invariance under  $\exp(tX)$ ,  $0 \leq t \leq 2$ . Finally, implicit constants in the notation  $\ll$  are permitted to depend on  $d$ , if a Sobolev norm  $\mathcal{S}_d$  is implicitly present.

The primary purpose of this section is to prove the following (“Proposition D” of §4).

**8.1. Proposition.** *Fix a Sobolev norm  $\mathcal{S}_d$ ; all notions of almost invariance are taken with respect to this. Suppose that  $\mu$  is  $\epsilon$ -almost invariant under a subgroup  $S$ , and also under  $Z \in \exp(\mathfrak{r})$ , where  $\|Z\| = 1$ .*

*Then there is a constant  $\kappa_{13}$  so that  $\mu$  is also  $c_1(d)\epsilon^{\kappa_{13}}$ -almost invariant under some subgroup  $S_*$  with  $\dim(S_*) > \dim(S)$ . If  $H \subseteq S$ , we may also assume that  $H \subseteq S_*$ .*

The proof could be much simplified, in our present setting, by using the assumption that there are finitely many intermediate subgroups between  $H$  and  $G$ . However, we shall make use of our (quite general) results from §7 on effective generation of Lie algebras.

**8.2. Stability properties of almost invariance.** The notion of “ $\mu$  is almost invariant under  $Z$ ,” for  $Z \in \mathfrak{g}$ , is almost stable under linear combinations, under commutators, and under passing from  $Z$  to a nearby element  $Z'$ :

**8.2.1. Lemma.** *let  $k \geq 1$ ,  $0 < \delta < 1$ . Let  $T \subset \mathfrak{g}$  consist of unit vectors, i.e.  $\|X\| = 1$  for  $X \in T$ . Let  $Z \in \mathfrak{g}$  be a unit vector so that  $\|Z - \sum_{t \in T^{(k)}} c_t t\| \leq \delta$ .*

*If  $\mu$  is  $\epsilon$ -almost invariant under every  $X \in T$  w.r.t.  $\mathcal{S}_d$ , it is also  $\ll (\epsilon^{\kappa_{21}^{(k)}} \sum_t |c_t| + \delta)$ -almost invariant under  $Z$  w.r.t.  $\mathcal{S}_d$ .*

*Proof.* The proof is a tedious elaboration of more or less obvious properties.

Let us observe

- If  $\mu$  is  $\epsilon$ -almost invariant under  $Z \in \mathfrak{g}$  with  $\|Z\| \leq 1$ , and  $W \in \mathfrak{g}$  satisfies  $\|W - Z\| \leq \delta$ , then  $\mu$  is  $\ll \max(\epsilon, \delta)$ -almost invariant under  $W$ .

This is an easy consequence of (3.13), (3.11) and the fact that  $\exp(tZ)$  and  $\exp(tW)$  are at distance  $\ll \delta$  from each other, for  $t \in [0, 2]$ . Next,

- If  $\mu$  is  $\epsilon$ -almost invariant under  $\exp(Z) \in \mathfrak{g}$ , and  $1 \leq c \leq 2\|Z\|^{-1}$ , then  $\mu$  is  $\ll (c\epsilon + \|Z\|)$ -almost invariant under  $cZ$ .

Indeed, it is evident that  $\mu$  is  $\ll n\epsilon$ -almost invariant under  $\exp(nZ)$ , when  $n$  is integral and satisfies  $n \leq 10\|Z\|^{-1}$ . This implies the stated conclusion.

- If  $\mu$  is  $\epsilon$ -almost invariant under  $Z \in \mathfrak{g}$ , and  $1 \leq c \leq \|Z\|^{-1}$ , then  $\mu$  is  $\ll c\epsilon$ -almost invariant under  $cZ$ .

This follows in a similar fashion to the previous statement.

- Suppose  $\mu$  is  $\epsilon$ -almost invariant under  $Z_1, Z_2 \in \mathfrak{g}$  with  $\|Z_i\| \leq 1$ . Take  $\alpha_1, \alpha_2$  so that  $|\alpha_i| \leq 1$  for  $i = 1, 2$ . Then  $\mu$  is  $\ll \epsilon^{1/2}$ -almost invariant under  $Z := \alpha_1 Z_1 + \alpha_2 Z_2$ .

We observe that, for  $n \geq 1$  integral,

$$\exp(\alpha_1 Z_1/n) \exp(\alpha_2 Z_2/n) = \exp\left(\frac{Z}{n}\right) \exp(W_1),$$

where  $\|W_1\| \ll 1/n^2$ . As a consequence, it follows that  $\mu$  is  $\ll (n^{-2} + \epsilon)$ -almost invariant under  $\exp(\frac{Z}{n})$ . It follows by a prior assertion that  $\mu$  is  $\ll (n^{-1} + n\epsilon)$ -almost invariant under  $Z$ . Take  $n = \epsilon^{-1/2}$  to conclude.

- Suppose  $\mu$  is  $\epsilon$ -almost invariant under  $Z_1, Z_2$  with  $\|Z_i\| \leq 1$ . Then  $\mu$  is  $\ll \epsilon^{1/2}$ -almost invariant under  $[Z_1, Z_2]$ .

Indeed,

$$\begin{aligned} \exp(Z_1/n) \exp(Z_2/n) \exp(Z_1/n)^{-1} \exp(Z_2/n)^{-1} = \\ \exp([Z_1, Z_2]/n^2) \exp(W_2), \end{aligned}$$

where  $\|W_2\| \ll n^{-3}$ . In view of that,  $\mu$  is  $\ll (\epsilon + n^{-3})$ -almost invariant under  $\exp([Z_1, Z_2]/n)$ . (Observe that our choice of the norm on  $\mathfrak{g}$  is so that  $\|[Z_1, Z_2]\| \leq 1$ .) By a previously noted fact,  $\mu$  is  $\ll (n\epsilon + n^{-2})$ -almost invariant under  $[Z_1, Z_2]$ . Take  $n = \epsilon^{-1/2}$  to conclude.

These remarks in hand, the assertion of the Lemma now follows easily. Regard  $k$  as fixed. We first note that  $\mu$  is  $\epsilon^{\kappa(k)}$ -almost invariant under each  $t \in T^{(k)}$ . (Observe also that  $\|t\| \leq 1$  for every such  $t$ ). From this, we deduce that  $\mu$  is also  $\epsilon^{\kappa'(k)}$ -almost invariant under  $C^{-1} \sum_t c_t t$  where  $C = \sum_t |c_t|$ . Thus  $\mu$  is  $\ll (C\epsilon^{\kappa_{21}(k)} + \delta)$ -almost invariant under  $Z$ , as required.  $\square$

**8.2.2. Lemma.** *There is a constant  $\kappa_{22}$  so that:*

*Let  $Q = S^0$  be the connected component of any intermediate subgroup  $H \subset S \subset G$ ; let  $\mu$  be  $\epsilon$ -almost invariant under  $Q$  w.r.t.  $\mathcal{S}_d$ . Then*

$$|\mu^q(f) - \mu(f)| \ll \epsilon \|q\|^{d\kappa_{22}} \mathcal{S}_d(f).$$

*Proof.* Because there are only finitely many possibilities for  $Q$ , it suffices to prove it for each possible  $Q$  individually. (This marks a point where we make usage of  $C_{\mathfrak{g}}(\mathfrak{h}) = \{0\}$ ). We claim that there are constants  $c_1, c_2$  depending only on  $Q$  so that, for every  $r \geq 2$ :

(8.1) Every  $q \in Q$  with  $\|q\| \leq r$  may be expressed as the product

$$\text{of } \leq c_1 + c_2 \log r \text{ elements with } \|q\| \leq 2$$

This follows from the structure theory of semisimple groups:  $Q$  is the connected component of the real points of a semisimple algebraic group. So there exists a compact subgroup  $K_Q \subset Q$  and a Cartan subalgebra  $\mathfrak{a}_Q \subset \mathfrak{q}$

so that  $Q = K_Q \exp(\mathfrak{a}_Q) K_Q$ . We are reduced to verifying (8.1) for elements of  $\exp(\mathfrak{a}_Q)$ , which is elementary.

Let  $q \in Q$ . By (8.1) there exist  $q_1, \dots, q_n \in Q$  with  $q = q_1 \cdots q_n$ ,  $\|q_i\| \leq 2$  and  $n \leq c_1 + c_2 \log \|q\|$ . Let  $\bar{q}_j = q_{j+1} \cdots q_n$  for all  $j = 1, \dots, n-1$ ,  $\bar{q}_n = 1$ . We have

$$(8.2) \quad |\mu^q(f) - \mu(f)| = \left| \sum_{j=1}^n \mu^{q_j}(\bar{q}_j \cdot f) - \mu(\bar{q}_j \cdot f) \right| \leq \epsilon \sum_{j=1}^n \mathcal{S}_d(\bar{q}_j \cdot f) \\ \ll_d \epsilon(1 + \log \|q\|) \mathcal{S}_d(f) (2N)^{\kappa_6 n d} \ll \epsilon \|q\|^{d\kappa_{22}} \mathcal{S}_d(f).$$

by our assumption, (3.11), and (3.3). As in (3.3) the integer  $N$  gives the dimension of the general linear group (with respect to which we defined  $\|q\|$ ).  $\square$

**8.3. Proof of Proposition 8.1.** It should be noted that this proof could be very considerably simplified by using the fact that there are finitely many intermediate subalgebras between  $\mathfrak{h}$  and  $\mathfrak{g}$ . However, the proof we indicate will work in a more general setup.<sup>23</sup>

We observe we are free to assume that  $\epsilon$  is *sufficiently small*, i.e.  $\epsilon \leq \epsilon_0(G, H)$ , in the statement of the Proposition. If not, the statement is obvious.

Assumptions as in the statement of the Proposition 8.1. Define  $T$  to be a finite subset of  $\mathfrak{g}$  obtained by adjoining  $Z$  to an arbitrary, orthonormal basis for  $\mathfrak{s}$ . Let  $k, c$  be as in Proposition 7.1. We apply Proposition 7.1 in the form indicated in its last sentence, with  $\delta = \epsilon^\alpha$ , where  $\alpha = \frac{\kappa_{21}}{k+c}$ . It produces a subalgebra  $\mathfrak{w} \supset \mathfrak{h}$ , with basis  $w_i$ .

By part (2) of Proposition 7.1,  $T$  is “almost contained in  $\mathfrak{w}$ ,” this implies, in particular, that  $\dim(\mathfrak{w}) > \dim(\mathfrak{s})$  so long as  $\epsilon$  is sufficiently small.

By part (1) of Proposition 7.1 and Lemma 8.2.1,  $\mu$  is  $\ll \epsilon^{c\alpha}$ -almost invariant under each  $w_i$ . From this, one deduces easily that  $\mu$  is also  $\epsilon^*$ -almost invariant under  $W$ , the connected Lie group with Lie algebra  $\mathfrak{w}$ . By our construction,  $W \supset H$  and  $\dim(W) > \dim(S)$ .  $\square$

## 9. EFFECTIVE ERGODIC THEOREM.

In this section, we establish an effective version of the Birkhoff ergodic theorem, i.e. we prove “Proposition A” from §4. Prior to doing so, we give a precise quantification of the notion of “generic” we shall need.

Throughout,  $\mu$  will denote an  $H$ -invariant  $H$ -ergodic measure on  $X = \Gamma \backslash G$ , so that  $L_0^2(\mu)$  is  $\frac{1}{p_G - 1}$  tempered as an  $H$ -representation. This applies, in particular, to the measures of central interest in this paper:  $\mu$  as in the

<sup>23</sup>More precisely, the proof written establishes that, under the assumptions of the Proposition 8.1, the measure  $\mu$  is almost invariant under an intermediate Lie subalgebra  $\mathfrak{h} \subset \mathfrak{q} \subset \mathfrak{g}$ . This statement makes no use of the finiteness of intermediate subgroups. However, in passing from the Lie algebra  $\mathfrak{q}$  to the group, we invoke Lemma 8.2.2, which does use the finiteness assumption.

statement of Theorem 1.3; or also to the  $S$ -invariant probability measures on connected closed  $S$ -orbits, for  $H \subset S \subset G$ . These comments, and the definition of  $p_G$ , follow from Lemma 6.7.1, together with the fact that the restriction of a tempered  $S$ -representation to  $H$  remains tempered.

**9.1. Generic points.** Let  $T \geq 1$ . Let  $M = 20(1 + p_G)$ , where  $p_G$  was defined in Lemma 6.7.1.<sup>24</sup> We define  $D_T(f)$  as the discrepancy between a horocycle average of  $f$  over a big stretch of the orbit and its integral:

$$(9.1) \quad D_T(f)(x) := \frac{1}{(T+1)^M - T^M} \int_{T^M}^{(T+1)^M} f(xu(t)) dt - \int f d\mu.$$

Clearly,  $D_T$  depends on the choice of the unipotent subgroup. Since we regard it as fixed, we suppress that dependence.

We say a point  $x \in X$  is  $T_0$ -generic w.r.t. the Sobolev norm  $\mathcal{S}$  if, for all integers  $n \geq T_0$  and all  $f \in C^\infty(X)$  we have the bound

$$(9.2) \quad |D_n(f)(x)| \leq n^{-1} \mathcal{S}(f).$$

We say a point  $x \in X$  is  $[T_0, T_1]$ -generic w.r.t.  $\mathcal{S}$  if the bound (9.2) holds for all integers  $n \in [T_0, T_1]$ .

**9.2. Proposition.** *For  $d \geq d_0$ , where  $d_0$  depends only on  $G, H$ , the measure of points that are not  $T_0$ -generic w.r.t.  $\mathcal{S}_{d_0}$  is  $\ll T_0^{-1}$ .*

*Proof.* First, consider a fixed  $f \in L^2(X)$  which is in the completion of  $C_c^\infty(X)$  with respect to  $\mathcal{S}_d$  (for a  $d$  specified below). The decay of matrix coefficients (cf. §6.3, esp. (6.4), (6.9) and) implies that:

$$\left| \langle u(t)f, f \rangle - \left( \int f d\mu \right)^2 \right| \ll (1 + |t|)^{-1/p_G} \mathcal{S}_{\dim H}(f)^2.$$

Let  $\square$  denote the square  $[T^M, (T+1)^M]^2 \subset \mathbb{R}^2$ . By definition of  $D_T(f)$ ,

$$\int_X |D_T(f)|^2 d\mu = \frac{\int_{\square} \langle u(t)f, u(s)f \rangle dt ds}{((T+1)^M - T^M)^2} - \left( \int f d\mu \right)^2.$$

Split  $\square$  into the sets where  $|s - t| \leq T^{\frac{M}{2}}$ , and where  $|s - t| \geq T^{\frac{M}{2}}$ . Thus:

$$\begin{aligned} \int_X |D_T(f)|^2 d\mu &\ll \frac{1}{T^{2M-2}} (T^{M-1+\frac{M}{2}} + T^{2M-2-\frac{M}{2p_G}}) \mathcal{S}_{\dim H}(f)^2 \\ &\ll T^{-4} \mathcal{S}_{\dim H}(f)^2 \end{aligned}$$

In the last equality, we have used the fact that  $M \geq 10 + 10p_G$ .

Therefore,

$$(9.3) \quad \mu(\{x : |D_T(f)(x)| \geq s\}) \ll s^{-2} T^{-4} \mathcal{S}_{\dim H}(f)^2$$

for any  $s > 0$ .

<sup>24</sup>In practice, it needs just to be a sufficiently large fixed number in our arguments; we have used the notation  $M$  to avoid distracting the reader with its specific value, which is irrelevant.

To obtain the lemma for all functions  $f \in C_c^\infty(X)$  we begin by choosing  $d > d' > \dim H$  so that the relative traces  $\text{Tr}(\mathcal{S}_{d'}^2 | \mathcal{S}_d^2)$  and  $\text{Tr}(\mathcal{S}_{\dim H}^2 | \mathcal{S}_{d'}^2)$  are both finite, cf. (3.10) and §5.2.

Next, choose an orthonormal basis  $e^{(1)}, \dots, e^{(r)}, \dots$  for the completion of  $C_c^\infty(X)$  with respect to the Hermitian norm  $\mathcal{S}_d$ . By the spectral theorem, we may choose such a basis which is also orthogonal w.r.t.  $\mathcal{S}_{d'}$ . Note that  $\mathcal{S}_{\dim H}$  is continuous with respect to the  $\mathcal{S}_{d'}$ , so that  $\sum_i \left( \frac{\mathcal{S}_{\dim H}(e^{(i)})}{\mathcal{S}_{d'}(e^{(i)})} \right)^2 < \infty$  by our assumption on the relative trace. We understand the summand as zero when  $\mathcal{S}_{d'}(e^{(i)}) = \mathcal{S}_{\dim H}(e^{(i)}) = 0$ .

Let  $c > 0$  be a constant, which we will specify below in a way depending only on  $X$ . Then, by applying (9.3) with  $s = cn^{-1}\mathcal{S}_{d'}(e^{(k)})$  and  $T = n$  for each  $n \geq T_0$  and for each  $k \geq 1$  we have:

$$\mu \left( \bigcup_{n \geq T_0, k \geq 1} \{x : n|D_n(e^{(k)})(x)| \geq c\mathcal{S}_{d'}(e^{(k)})\} \right) \ll T_0^{-1}c^{-2} \ll T_0^{-1}.$$

Let  $E$  be the set indicated on the left-hand side. We claim that any  $x \notin E$  is generic w.r.t.  $\mathcal{S}_d$  (once  $c$  has been chosen correctly).

Take  $f = \sum_k f_k e^{(k)} \in C_c^\infty(X)$ . We have for any  $n \geq T_0$ , and any  $x \notin E$ ,

$$\begin{aligned} n|D_n(f)(x)| &\leq c \sum_k |f_k| \mathcal{S}_{d'}(e^{(k)}) \leq \\ &c \left( \sum_k |f_k|^2 \right)^{1/2} \left( \sum_k \mathcal{S}_{d'}(e^{(k)})^2 \right)^{1/2} = c\mathcal{S}_d(f) \left( \sum_k \mathcal{S}_{d'}(e^{(k)})^2 \right)^{1/2} \end{aligned}$$

by linearity of  $D_n(\cdot)$  and the finiteness of the relative trace (3.10). Choosing  $c$  equal to the inverse of the last square root, we have shown  $|D_n(f)(x)| \leq n^{-1}\mathcal{S}_d(f)$  for  $n \geq T_0$  and  $x \notin E$  as required.  $\square$

We require also an adaptation of this Proposition to the setting when one works with a measure  $\mu$  that is almost invariant (“Proposition A”). We refer to §4.3 for a discussion of its meaning in qualitative terms.

**9.3. Proposition.** *Let  $H \subset S \subset G$ ,  $S$  connected. Suppose that  $\mu$  is  $\epsilon$ -almost invariant under  $S$  w.r.t.  $\mathcal{S}_d$ , for  $d \geq \kappa_5 + 1$ . Then there exists  $\beta \in (0, 1/2)$ ,  $d' > d$ , depending only on  $G$ ,  $H$ , and  $d$ , so that:*

*Whenever  $R \leq \epsilon^{-\beta}$  and  $T_0 > 0$ , the fraction of points  $(x, s) \in X \times B_S(R)$  w.r.t.  $\mu$  resp. the Haar measure on  $S$  for which  $x.s$  is not  $[T_0, \epsilon^{-\beta}]$ -generic w.r.t.  $\mathcal{S}_{d'}$  is  $\ll_d T_0^{-1}$ .*

*Proof.* Let  $B := B_S(R)$ . As in the proof of Proposition 9.2 we will first estimate

$$\frac{1}{\text{vol}(B)} \int_{X \times B} |D_T(f)(xs)|^2 d\mu ds.$$

This equals

$$\frac{1}{\text{vol}(B)} \int_B \int_X |D_T(f)|^2 d\mu^s(x) ds \ll \epsilon R^{d\kappa_{22}} \mathcal{S}_d(|D_T(f)|^2) + \int_X |D_T(f)|^2 d\mu$$

where  $\mu^s$  denotes the translated measure; we have used the definition of *almost invariant* from §3.10, the fact that  $B_S(R)$  consists of elements with  $\|s\| \leq R$ , and Lemma 8.2.2.

As in the proof of Proposition 9.2, the latter term is bounded, up to a constant depending only on  $G$  and  $H$  by  $T^{-4} \mathcal{S}_{\dim(H)}(f)^2$ .

On the other hand, there exists a constant  $\kappa$  so that the first term is, by (3.11) and (3.12),

$$\ll \epsilon R^{d\kappa_{22}} T^{d\kappa} \mathcal{S}_{d+\kappa_5}(f)^2.$$

We take  $\beta = \frac{1}{2}(4 + \kappa_{22}d + \kappa d)^{-1}$ . This choice is made so that

$$\epsilon R^{d\kappa_{22}} T^{d\kappa} \leq T^{-4},$$

whenever  $T \leq \epsilon^{-\beta}$ ,  $R \leq \epsilon^{-\beta}$ .

Therefore, if  $d \geq \dim(H)$ ,  $T \leq \epsilon^{-\beta}$ ,  $R \leq \epsilon^{-\beta}$ ,

$$\frac{1}{\text{vol}(B)} \int_B \int_X |D_T(f)|^2 d\mu^s(x) \ll T^{-4} \mathcal{S}_{d+\kappa_5}(f)^2.$$

Reasoning as in the second part of the proof of Proposition 9.2, and increasing  $d$ , yields the conclusion for all  $f \in C_c^\infty(X)$ .  $\square$

## 10. NEARBY GENERIC POINTS EFFECTIVELY GIVE ADDITIONAL INVARIANCE.

In this section, we shall use the polynomial properties of unipotent flows and the effective ergodic theorem to establish “Proposition B.”

Henceforth,  $\mathfrak{r}$  will denote an  $S$ -invariant complement to the Lie algebra of  $S$ , inside  $\mathfrak{g}$ . It exists because  $S$  is semisimple. We recall we have fixed a unipotent one-parameter subgroup  $u(t) \subset H$ , see (3.1).

There exists some  $\kappa_{23} > 0$  with the following property: whenever  $v, w \in \mathfrak{r}$   $\|v\|, \|w\| \leq \kappa_{23}$ ,

$$(10.1) \quad \exp(v) \exp(w)^{-1} = \exp(w^*)s, w^* \in \mathfrak{r}, s \in S.$$

where  $d(s, 1) \leq \|v - w\|$  and  $\frac{\|w^*\|}{\|v - w\|} \in [1/2, 2]$ . Indeed, the map  $(v, w) \mapsto (w^*, s)$  is differentiable; also  $(v, v) \mapsto (0, e)$  and the derivative at  $(0, 0)$  is the map  $(v, w) \mapsto (v - w, 0)$ . The assertion follows.

**10.1. Unipotent trajectories of nearby points.** The discussion that follows may be regarded as an effective form of the “equicontinuity” statements from §2.5.

Let  $x_1, x_2 \in X$  so that  $x_2 = x_1 \exp(r)$ , some  $r \in \mathfrak{r}$ . Note that  $x_2 u(t) = x_1 u(t) \exp(\text{Ad}(u(-t))r)$ . We decompose  $r = r_0 + r_1$  according to (3.2).

Then  $\text{Ad}(u(-t))r_0 = r_0$  for all  $t$ , whereas  $t \mapsto \text{Ad}(u(-t))r_1$  is a polynomial of degree  $\leq \dim(\mathfrak{g})$  and all of whose coefficients (w.r.t. an orthonormal basis for  $\mathfrak{g}$ ) are  $\ll \|r_1\|$ .

Suppose  $r_1 \neq 0$ . Then there exists a positive real time  $T$  satisfying  $\|r\|^{-*} < T < \|r_1\|^{-*}$ , and a polynomial  $q : \mathbb{R} \rightarrow \mathfrak{r}$  of degree  $\leq \dim(\mathfrak{g})$  with image centralized by  $u(\mathbb{R})$  satisfying  $q(0) = 0$  and  $\max_{s \in [0, 2]} \|q(s)\| = \kappa_{23}$ , so that:

$$(10.2) \quad \text{Ad}(u(-t)).r = q(t/T) + O(\|r_1\|^*) \text{ for all } t \leq 3T.$$

Here, and in the following, the implicit constant in the  $O(\cdot)$ -notation is understood in the same sense as in the  $\ll$ -notation.

If  $r_1 = 0$ , the same statement remains true, except one sets “ $T = \infty$ ” and ignores the statement about  $\max \|q(s)\|$ ; the polynomial  $q$  is in this case constant.

This statement will be used to give an effective version (“Proposition B”) of (2.2).

**10.2. Proposition.** *Let  $d \geq \kappa_5 + 1$ . There exists constants  $\kappa_7 > 0$  and  $\kappa_8 > \kappa_9 > 0$  so that:*

*Suppose  $\mu_1, \mu_2$  are  $H$ -invariant measures, that  $x_1, x_2 \in X$  satisfy  $x_2 = x_1 \exp(r)$  for some nonzero  $r \in \mathfrak{r}$ , and that  $x_i$  is  $[\|r\|^{-\kappa_8}, \|r_1\|^{-\kappa_9}]$ -generic w.r.t.  $\mu_i$  and a Sobolev norm  $\mathcal{S}_d$  (for  $i = 1, 2$ ). Then there is a polynomial  $q : \mathbb{R} \rightarrow \mathfrak{r}$  of degree  $\leq \dim(\mathfrak{g})$ , so that:*

$$(10.3) \quad \left| \mu_2^{\exp q(s)}(f) - \mu_1(f) \right| \ll_d \|r_1\|^{\kappa_7} \mathcal{S}_d(f), 1 \leq s \leq 2^{1/M}$$

*If  $r_1 \neq 0$ , then  $\max_{s \in [0, 2]} \|q(s)\| = \kappa_{23}$ . Moreover, if  $\mu_1 = \mu_2$  is  $\epsilon$ -almost invariant under  $S$ , then  $\mu_1$  is  $\ll_d \max(\epsilon, \|r\|^{\kappa_7})^{1/2}$ -almost invariant under some  $Z \in \mathfrak{r}$  with  $\|Z\| = 1$ .*

*Proof.* We let  $T$  be as in the discussion before the proposition. The definition of generic (§9.1) assures us (assuming that  $\kappa_9$  is sufficiently small and  $\kappa_8$  is sufficiently large) for any integer

$$n \in [T^{1/M}, (2T)^{1/M}] \subset [\|r\|^{-\kappa_9}, \|r_1\|^{-\kappa_8}]$$

and for any  $f \in C_c^\infty(X)$ , we have

$$\left| \int f d\mu_j - \frac{1}{(n+1)^M - n^M} \int_{n^M}^{(n+1)^M} f(x_j u(t)) dt \right| \leq n^{-1} \mathcal{S}_d(f),$$

for  $j = 1, 2$ .

For  $t, t_0 \in [n^M, (n+1)^M]$ , we have, using (3.13) and (10.2)

$$\begin{aligned} f(x_2 u(t)) &= f(x_1 u(t) \exp(q(t/T))) + O(\|r_1\|^* \mathcal{S}_d(f)) \\ &= f(x_1 u(t) \exp q(t_0/T)) + O(T^{-1/M} \mathcal{S}_d(f)) + O(\|r_1\|^* \mathcal{S}_d(f)). \end{aligned}$$

Here we used in the last line that  $|t - t_0| \ll n^{M-1} \asymp T^{1-1/M}$  for  $t, t_0 \in [n^M, (n+1)^M]$  — which shows that the polynomial  $q(t/T)$  has small variation within that interval.

Thus,

$$(10.4) \quad \mu_2(f) = \mu_1(\exp q(t_0/T).f) + O\left((T^{-1/M} + \|r_1\|^*)\mathcal{S}_d(f)\right)$$

which implies (10.3).

Let us observe that if  $r_1 = 0$ , then the polynomial  $q$  is constant, say  $q \equiv q_0$ , and we have  $\mu_2(f) = \mu_1(\exp(q_0)f)$  (exactly).

Assume in the remainder of the proof that  $\mu_1 = \mu_2$ . The prior remark shows that, if  $r_1 = 0$ ,  $\mu$  is indeed  $\|r\|$ -almost invariant under  $Z = \|r\|^{-1}r$  by (3.13); this establishes the final assertion in the case when  $r_1 = 0$ .

Assume, thus, that  $r_1 \neq 0$ . In this case  $q(0) = r$  and  $\max_{s \in [0,2]} \|q(s)\| = \kappa_{23}$ , so that the coefficients of  $q'(s)$  are  $\gg 1$ . By (10.3),  $\mu_1$  is  $\ll \|r\|^{\kappa_7}$ -almost invariant under  $\exp(q(s))$  for any  $s \in [1, 2^{1/M}]$ .

Put  $E = \max(\epsilon, \|r\|^{\kappa_7})^{1/2}$ . There exists  $s \in [1, 2^{1/M}]$  with  $s+E \in [1, 2^{1/M}]$  for which  $\|q(s) - q(s+E)\| \gg E$ . (Rather, we may assume this is so; if  $E$  is so large that this is false for a trivial reason, then the final statement of the Proposition correspondingly becomes trivial).

With  $s$  being so chosen,  $\mu_1$  is  $\ll \|r\|^{\kappa_7}$ -almost invariant under  $\exp(q(s))$ ,  $\exp(q(s+E))$ , and so also under  $\exp(-q(s))\exp(q(s+E))$ . Put  $v = q(s)$ ,  $w = q(s+E) - v$ . Then, by (10.1),  $\exp(-v)\exp(v+w) = \exp(w^*)s$  where  $w^* \in \mathfrak{r}$ ,  $\|w^*\| \asymp E$ , and  $s \in S$  satisfies  $d(s, e) \ll E$ .

This claim proves the proposition: Since  $\mu_1$  is assumed to be  $\epsilon$ -almost invariant under  $S$ , we get that  $\mu$  is  $\ll \max(\epsilon, \|r\|^{\kappa_7}) = E^2$ -invariant under  $\exp(w^*)$ . Since  $w \in \mathfrak{r}$  with  $\|w^*\| \asymp E$  we can iterate this statement  $E^{-1}$ -many times; using (3.13), we deduce that  $\mu$  is  $\ll E$ -almost invariant under  $Z = \frac{1}{\|w^*\|}w^*$ . See the proof of Lemma 8.2.1 (esp. second bulleted point) for details of this type of iteration.

□

**10.3. Quantitative isolation of closed orbits.** From this argument we may draw the following useful corollary regarding the isolation of closed orbits for semisimple groups. It gives a quantitative estimate on the spacing between two distinct, closed  $S$ -orbits, as well on how closely such an orbit can approach itself.

The idea of proof is simple. Suppose given two very nearby points  $x_1, x_2$  so that  $x_1S, x_2S$  are both closed and are distinct. By modifying  $x_1, x_2$  slightly, we may assume that we are in the situation of Proposition 10.2, with  $\mu_i$  the  $S$ -invariant probability measure on  $x_iS$ . The conclusion of Proposition 10.2 then implies that  $x_2S$  contains, loosely speaking, “many different translates of  $x_1S$  along the direction  $q(s)$ .” Thus, the volume of  $x_2S$  was necessarily large.

**10.3.1. Lemma.** *There are constants  $\kappa_{10}$  and  $\kappa_{11}$  with the following property.*

*Suppose  $H \subset S \subset G$ . Let  $x_1, x_2 \in X$  be so that  $x_iS$  are closed orbits with volume  $\leq V$ . Then either:*

- (1)  $x_1$  and  $x_2$  are on the local  $S$ -orbit, i.e. there exists some  $s \in S$  with  $d(s, 1) \leq 1$  and  $x_2 = x_1 s$ , or
- (2)  $d(x_1, x_2) \gg \min(\text{ht}(x_1), \text{ht}(x_2))^{-\kappa_{10}} V^{-\kappa_{11}}$ .

The proof will use a simple argument, which we will also use elsewhere. Recalling the definition of  $r_1$  from (3.2), there exists  $\iota_6 > 0$  so that, for any  $r \in \mathfrak{r}$ ,

$$(10.5) \quad \text{vol}\left\{s : d(s, 1) \leq 1 \text{ with } \frac{\|(\text{Ad}(s^{-1})r)_1\|}{\|r\|} \leq \iota_6\right\} < \frac{\text{vol}(\{s \in S : d(s, 1) \leq 1\})}{2}.$$

Indeed, given  $r' \in \mathfrak{r}$ , there exists  $\varepsilon(r') > 0$  so that the measure of  $\left\{s : d(s, 1) \leq 1 \text{ with } \frac{\|(\text{Ad}(s^{-1})r')_1\|}{\|r'\|} \leq \varepsilon\right\}$  is less than  $\frac{1}{2} \text{vol}(\{s \in S : d(s, 1) \leq 1\})$ . This is so, since for the function  $s \mapsto (\text{Ad}(s^{-1})r')_1$  is real-analytic and not identically zero; if it were 0, then  $r'$  is centralized by every conjugate of  $\{u(t)\}$ , and so also by the Lie algebra of  $S$ , a contradiction. By a compactness argument, one may choose  $\varepsilon = \iota_6$  uniform over  $r' \in \mathfrak{g}$ , establishing (10.5).

*Proof.* Throughout, we fix a Sobolev norm  $\mathcal{S}_d$ , where  $d$  is sufficiently large (depending only on  $G, H$ ).<sup>25</sup>

We may suppose (adjusting  $x_1$  or  $x_2$  by an element of  $S$ ) that  $x_1 = x_2 \exp(r)$ , with  $r \in \mathfrak{r}$ , and will establish a lower bound on  $\|r\|$ .

Let  $\mu_i$  be the  $S$ -invariant probability measure on  $x_i S$ . In view of Proposition 9.2, the measure of

$$\{s \in S : d(s, 1) \leq 1 \text{ and } x_i s \text{ fails to be } V^2\text{-generic w.r.t. } \mu_i\},$$

with respect to the volume measure on  $S$ , is  $\ll V^{-1}$  for  $i \in \{1, 2\}$ . We may suppose that  $V$  is sufficiently large, so that this measure, as a fraction of the volume of  $\{s \in S : d(s, 1) \leq 1\}$ , is at most  $\frac{1}{10}$ .

We claim there exists  $s \in S$ , with  $d(s, 1) \leq 1$  with:

$$(10.6) \quad x_i s \text{ are } V^2\text{-generic w.r.t. } \mu_i \text{ and } \|(\text{Ad}(s^{-1})r)_1\| \geq \iota_6 \|\text{Ad}(s^{-1})r\|.$$

Indeed, let  $\Upsilon \subset \{s \in S : d(s, 1) \leq 1\}$  consist of those  $s$  for which  $x_i s$  is  $[V^2, \infty)$ -generic w.r.t.  $\mu_i$ . By (10.5),  $\Upsilon$  must contain an element  $s$  satisfying (10.6).

Replace  $x_i$  by  $x_i s$  and  $r$  by  $\text{Ad}(s^{-1})r$ , where  $s$  is so that (10.6) is satisfied. This has the effect of increasing  $\|r\|$  by, at most, a constant factor depending only on  $G, H$ . Thereby, we may assume that  $x_i$  are both  $V^2$ -generic and that  $\|r_1\| \gg \|r\|$ .

Let  $\eta > 0$  be smaller than the injectivity radius  $\iota_4 \text{ht}(x)^{-\kappa_3}$  of (3.7) (specified below). Let  $f$  be a smooth bump function on  $G$  supported in  $\{g : d(g, 1) \leq \eta\}$  with  $0 \leq f \leq 1$  which is one in a neighborhood of radius  $\gg \eta$  and whose partial derivatives up to order  $d$  are everywhere  $\ll \eta^{-d}$ .

<sup>25</sup> Precisely,  $d \geq \max(d_0, \kappa_5 + 1)$ ,  $d_0$  as in Lemma 9.2.

Let  $f_{x_1}$  be the function defined by  $x_1 g \mapsto f(g)$  when  $d(g, 1) \leq \eta$ , and zero outside that ball.

Then, evidently,  $\int f_{x_1} d\mu_1 \gg V^{-1} \eta^{\dim S}$ . On the other hand, the definitions of the Sobolev norm imply that

$$(10.7) \quad \mathcal{S}_d(f_{x_1}) \ll \text{ht}(x_1)^d \eta^{-d}.$$

Without loss of generality,  $\|r\| \leq V^{-2/\kappa_8}$ . We apply Proposition 10.2. Therefore, Proposition 10.2 assures us that, so long as

$$(10.8) \quad \text{ht}(x_1)^d \eta^{-d} \|r\|^{\kappa_7} \leq c V^{-1} \eta^{\dim(S)}$$

(for some  $c$  obtained from the above implicit constants), then,

$$(10.9) \quad \int f_{x_1}^{q(s)} d\mu_2 > 0 \text{ for } s \in (1, 2^{1/M}).$$

In words, (10.9) asserts that the orbit  $x_2 S$  passes very close to  $x_1 S$ , many times. More precisely, given  $s_1, s_2$  so that  $\eta \ll d(q(s_1)^{-1}, q(s_2)^{-1}) \ll \text{ht}(x)^{-\kappa_3}$ , the functions  $f_{x_1}^{q(s_1)}$  and  $f_{x_1}^{q(s_2)}$  have distinct support. Moreover,  $x_2 S$  must intersect both of their supports. Thereby, we obtain from  $s_1, s_2$ , two distinct discs of radius  $\asymp \text{ht}(x)^{-\kappa_3}$  on  $x_2 S$ .

Using not just  $s_1, s_2$  but many such  $s_i$ s, we conclude that:

$$(10.10) \quad \text{vol}(x_2 S) \gg \text{ht}(x_1)^{-\kappa_3(\dim S + 1)} \eta^{-1}.$$

Take  $\eta$  so that the right hand side is (including the implicit constant) equal to  $2V$ . Then (10.8) cannot hold, for  $\text{vol}(x_2 S) \leq V$ . This bounds  $\|r\|$  from below by an expression of the form  $\text{ht}(x_1)^{-\star} V^{-\star}$  as required.  $\square$

## 11. THE MEASURE OF POINTS NEAR PERIODIC ORBITS IS SMALL.

Let  $\mu$  be as in Theorem 1.3. Using the polynomial properties of unipotent flows, together with the quantitative isolation of periodic orbits established in Lemma 10.3.1, we will be able to establish that the total  $\mu$ -measure of the set of points near periodic  $S$ -orbits of bounded volume remains rather small.

The proof proceeds as follows. By Lemma 10.3.1, we can restrict attention to a single periodic  $S$ -orbit. We then divide  $X$  into a piece near the cusp, and its complement. The measure of the piece near the cusp is handled by Lemma 3.6.1; the complement is handled using a linearization of the flow near the periodic  $S$ -orbit.

**11.1. Proposition.** *Let  $S \supset H$ . There exists some  $V_0$  depending on  $\Gamma$ ,  $G$ , and  $H$  and some  $\kappa_{12} > 0$  with the following property. Let  $V \geq V_0$  and suppose  $\mu(Y) = 0$  if  $Y$  is any closed  $S$ -orbit of volume  $\leq V$ . Then:*

$$(11.1) \quad \mu(\{x \in X : \text{there exists } x' \stackrel{V^{-\kappa_{12}}}{\sim} x \text{ with } \text{vol}(x' S) \leq V\}) \leq 1/2$$

*Proof.* This is a quantitative form of polynomial nondivergence. More precisely, we use Lemma 10.3.1 together with the “linearization” technique (see §1.7.1).

Let  $\mathfrak{r}$  be as in §10. Let  $x_0S$  be any closed  $S$ -orbit. In view of Lemma 10.3.1 and Lemma 3.6.1 (1), the number of such orbits with volume  $\leq V$  is bounded by a polynomial function of  $V$ .

We show that for any closed orbit  $x_0S$  of volume  $\leq V$ :

$$(11.2) \quad \mu((x_0S \cap \mathfrak{S}(R)) \exp\{r \in \mathfrak{r} : \|r\| \leq \delta\}) \ll V^* R^* \delta^*$$

We shall only sketch the argument for (11.2), for it is by now quite standard. The deduction of (11.1) is then straightforward: take  $R = V$ ,  $\delta$  the reciprocal of a large power of  $V$ , and apply Lemma 3.6.1,

Let  $\Omega_\delta = \{r \in \mathfrak{r} : \|r\| < \delta\}$ . For  $\delta_0 = R^{-*}V^{-*}$ , the map

$$(x_0S \cap \mathfrak{S}(R)) \times \Omega_{\delta_0} \rightarrow X, \quad (y, r) \mapsto y \exp(r)$$

is a diffeomorphism onto an open neighbourhood  $\mathcal{N}_{\delta_0}$  of  $x_0S \cap \mathfrak{S}(R)$ , as follows from the implicit function theorem, Lemma 10.3.1 and (3.7).

Let  $\pi : \mathcal{N} \rightarrow \Omega_{\delta_0}$  be the natural projection. For  $\delta \leq \delta_0$ , let  $\mathcal{N}_\delta = \pi^{-1}(\Omega_\delta)$ .

For  $y \in \mathcal{N}$  and  $t \in \mathbb{R}$ , we have:

$$(11.3) \quad \pi(yu(t)) = \text{Ad}(u(-t))\pi(y)$$

so long as  $yu(s) \in \mathcal{N}_{\delta_0}$  for all  $s \in [0, t]$ . Notice that the latter is equivalent to  $\text{Ad}(u(-s))\pi(y) \in \Omega_{\delta_0}$  for all  $s \in [0, t]$ .

Choose  $x \in X$  which is generic for the flow  $u(\cdot)$  w.r.t.  $\mu$ . Recall that, as in Section 2, this means the measure along the trajectory  $\{xu(t) : 0 \leq t \leq T\}$  approximates  $\mu$  as  $T \rightarrow \infty$ . Let us observe that, for any such  $x$  and all points  $y \in \{xu(t) \cap \mathcal{N}_{\delta_0}\}$ , the polynomial  $t \mapsto \text{Ad}(u(-t))\pi(y)$  is nonconstant. Suppose this to be false; set  $Y = \pi(y)$ . The genericity of  $y$  implies that  $\text{supp}(\mu) \subset x_0S \cdot \exp(Y)$ . Since  $\text{supp}(\mu)$  is an  $H$ -orbit, this implies  $H \subset \exp(-Y)S \exp(Y)$ . This is impossible, in view of Lemma 3.4.1, when  $\|Y\|$  sufficiently small. Since  $\|Y\| \leq \delta_0$  and  $\delta_0 = R^{-*}V^{-*}$ , we may assume  $\|Y\|$  is sufficiently small by changing the implicit constants.

Let  $\delta \leq \delta_0$ . Let  $Z_\delta = \{t \in \mathbb{R}_{\geq 0} : xu(t) \in \mathcal{N}_\delta\}$  be the set of times where the orbit is  $\delta$ -close to  $x_0S$ . Notice that  $Z_\delta$  is a union of intervals. There exists a constant  $\kappa_{24}$  so that we may cover  $Z_\delta$  by intervals  $B_j$  satisfying:  $\frac{|B_j \cap Z_\delta|}{|B_j|} \leq c_1 (\delta \delta_0^{-1})^{\kappa_{24}}$  where we use the notation  $|I|$  to denote the length of an interval  $I \subset \mathbb{R}$ . That follows from (11.3) and an argument using the growth properties of the polynomial  $t \mapsto \text{Ad}(u(-t))\pi(y)$ .<sup>26</sup> Here the intervals  $B_j$  can be chosen to be disjoint (since they correspond to different visits of the orbit of  $x$  to  $\mathcal{N}_0$ ).

We are using this to bound for a given  $T$  the fraction of times  $t \in [0, T]$  with  $xu(t) \in \mathcal{N}_\delta$ . Clearly, we may assume each  $B_j$  intersects  $[0, T]$ . Let  $J$

<sup>26</sup>Essentially, the argument here is that a polynomial which is  $\delta$ -small on an interval stays  $\delta_0$ -small on a bigger interval.

be the interval which is the union of  $[0, T]$  with the  $B_j$ . Then,

$$|J \cap Z_\delta| \leq \sum |B_j \cap Z_\delta| \leq c_2 (\delta \delta_0^{-1})^{\kappa_{24}} \sum |B_j| \ll (\delta \delta_0^{-1})^{\kappa_{24}} |J|.$$

This being true for a sequence of intervals  $J$  of increasing length, we see that

$$\mu(\mathcal{N}_\delta) \ll (\delta \delta_0^{-1})^{\kappa_{24}} \text{ whenever } \delta \leq \delta_0 = R^{-\star} V^{-\star}.$$

This is (11.2) (since for  $\delta \geq \delta_0$  the claim is trivial anyway).  $\square$

## 12. SOME LEMMAS CONNECTED TO LATTICE POINT COUNTING.

12.1. It is a general feature that, if  $H_1 \subset H_2$  are nice groups, the spectrum of  $L^2(H_2/H_1)$  is related to the volume growth of  $H_1$  inside  $H_2$ . In the case when  $H_1$  is a *lattice* inside  $H_2$ , this may be used to derive asymptotics for the number of points of  $H_1$  inside a large ball.

See [16, 19] for various instances of this technique in the context of Lie groups. A closely related idea, in a slightly different context, was introduced and utilized by G.M. in [41].

We shall need a slight variant, where we shall give upper bounds on the counting functions for *cosets* of a lattice.

**12.2. Proposition.** *Let  $H \subset S \subset G$  with  $S$  connected. Let  $\Lambda \subset S$  a discrete subgroup,  $B \subset S$  an open set,  $s_0 \in B$ .*

*Then:*

(1)

$$(12.1) \quad |\Lambda \cap \tilde{B}| \gg \frac{\text{vol}(B \cap S)}{\text{vol}(\Lambda \backslash S)} - C \int_{\tilde{B}} \varphi_0^{\text{wk}}(s)^\rho \, ds$$

(2) *the cardinality of any 1-separated subset  $\Delta \subset B \cap (\Lambda s_0)$  is bounded by*

$$(12.2) \quad |\Delta|^2 \ll \left( \frac{\text{vol}(\tilde{B})^2}{\text{vol}(\Lambda \backslash S)} + \int_{\tilde{B}} \varphi_0^{\text{wk}}(ss'^{-1})^\rho \, ds \, ds' \right)$$

*where:*

- $s \mapsto \varphi_0(s)$  is the Harish-Chandra spherical function associated to  $S$ , and  $\varphi_0^{\text{wk}}$  the modification for the case when  $S$  has multiple simple factors (see §6.3);
- $\rho > 0$  depends only on the spectral gap for  $L_0^2(\Lambda \backslash S)$  and  $C$  depends only on  $G, H$ . Here  $L_0^2$  denotes the orthogonal complement of locally constant functions.
- $\tilde{B} = \{s \in S : d(s, 1) \leq 1\} B \{s \in S : d(s, 1) \leq 1\}$ .

Moreover, we interpret  $\text{vol}(\Lambda \backslash S) := \infty$  if  $\Lambda$  is not of cofinite volume.

Note that one deduces from this bounds for the size of any  $\delta$ -separated subset of  $\Lambda s_0 \cap B$ , for  $\delta < 1$ . Indeed, if there exists a  $\delta$ -separated subset of a given metric space of size  $N_1$ , there exists a 1-separated subset of size

$\geq N_1/N_2$ , where  $N_2$  is the largest possible size of a  $\delta$ -separated set within a 1-ball.

*Proof.* Let  $\chi$  be a non-negative smooth function supported in the neighbourhood  $\Omega = \{s \in S : d(s, 1) \leq 1/2\}$ , satisfying  $\int \chi = 1$ . Note that  $\Omega = \Omega^{-1}$ . Set

$$B_1 = B.\Omega, B_2 = \Omega B_1 = \Omega.B.\Omega, B_3 = \Omega.B_2.\Omega,$$

and note that  $B_1, B_2, B_3$  are all contained in  $\tilde{B}$ .

Let  $\pi : S \rightarrow \Lambda \backslash S$  be the projection and  $\pi_* : C_c(S) \rightarrow C_c(\Lambda \backslash S)$  be the natural projection map, i.e.  $\pi_*(f)(\Lambda g) = \sum_{\lambda \in \Lambda} f(\lambda g)$ . Let  $\pi^*$  be the pullback  $C(\Lambda \backslash S) \rightarrow C(S)$ . Then, for  $f_1 \in C_c(S), f_2 \in C_c(\Lambda \backslash S)$ , we have:

$$\langle \pi_*(f_1), f_2 \rangle_{\Lambda \backslash S} = \langle f_1, \pi^* f_2 \rangle_S$$

Let  $f(s) = \int_{B_2} \chi(ss_1^{-1}) ds_1 \in C_c(S)$ . Then  $\text{supp}(f) \subset \Omega B_2$  and  $f \geq 1_{B_1}$ , where  $1_{B_1}$  is the characteristic function of  $B_1$ . Indeed, for  $s_1 \in B_1$  we have  $f(s_1) = \int_{B_2} \chi(s_1 s^{-1}) ds \geq \int_{\Omega s_1} \chi(s_1 s^{-1}) ds = 1$ . By these definitions and since  $0 \leq f \leq 1$  always, we have that

$$\begin{aligned} \langle \pi_* f, \pi_* \chi \rangle_{\Lambda \backslash S} &= \langle \pi^* \pi_* f, \chi \rangle_S \\ &= \sum_{\lambda \in \Lambda} \int_{s \in S} f(\lambda s) \chi(s) ds \leq |\Lambda \cap \text{supp}(f).\Omega| \leq |\Lambda \cap B_3|. \end{aligned}$$

Also note that  $\pi^* \pi_* 1_{B_1}(s) = \sum_{\lambda \in \Lambda} 1_{B_1}(\lambda s) = |\Lambda s \cap B_1|$  and so

$$\begin{aligned} (12.3) \quad \langle \pi_* f, \pi_* f \rangle_{\Lambda \backslash S} &= \langle f, \pi^* \pi_* f \rangle_S \geq \langle 1_{B_1}, \pi^* \pi_* 1_{B_1} \rangle_S = \\ &= \int_{B_1} |\Lambda s \cap B_1| ds = \int_{B_1} |\Lambda \cap B_1 s^{-1}| ds \end{aligned}$$

Next we need to understand the relationship between separated subsets of  $\Lambda s_0 \cap B$  and the expression  $|\Lambda \cap B_1 s^{-1}|$ . Suppose that there exists a 1-separated subset  $\Delta \subset \Lambda s_0 \cap B$ , for some  $s_0 \in B$ . Clearly for each  $\delta = \lambda s_0 \in \Delta$ , we have  $\Delta \delta^{-1} \subset \Lambda s_0 \delta^{-1} \cap B \delta^{-1} \subset \Lambda \cap B \delta^{-1}$ . In particular, whenever  $s \in \delta \Omega$ , we must have  $\Delta \delta^{-1} \subset \Lambda \cap B_1 s^{-1}$  and so  $|\Lambda \cap B_1 s^{-1}| \geq |\Delta|$ . Because the set  $\Delta$  is 1-separated, the balls  $\delta \Omega$  for  $\delta \in \Delta$  are disjoint. Each such  $\delta \Omega$  with  $\delta \in \Delta \subset B$  is contained in  $B \Omega \subset B_1$ . Therefore, the integral on the right hand side of (12.3) is  $\geq |\Delta|^2 \cdot \text{vol}(\Omega)$ .

We have shown that

$$|\Lambda \cap B_3| \geq \langle \pi_* f, \pi_* \chi \rangle \text{ and } |\Delta|^2 \leq \text{vol}(\Omega)^{-1} \|\pi_* f\|_{L^2(\Lambda \backslash S)}^2.$$

It remains to estimate  $\|\pi_* f\|_{L^2(\Lambda \backslash S)}$  and  $\langle \pi_* f, \pi_* \chi \rangle$ .

The integral of  $f$  equals  $\text{vol}(B_2)$ . Consequently, the projection of  $\pi_* f$  onto the locally constant functions has  $L^2$ -norm  $\asymp \frac{\text{vol}(B_2)}{\text{vol}(\Lambda \backslash S)^{1/2}}$ .

Similarly, if  $\text{Pr}$  denotes the orthogonal projection from  $L^2(\Lambda \backslash S)$  to constant functions, we see that  $\langle \text{Pr } \pi_* f, \text{Pr } \pi_* \chi \rangle \gg \frac{\text{vol}(B \cap S)}{\text{vol}(\Lambda \backslash S)}$ .

To handle the projection of  $\pi_* f$  onto the orthocomplement of the constants, we use standard bounds on matrix coefficients (see (6.7) – (6.9) and (6.10)). We write  $\text{Pr}_0$  for the projection onto the orthogonal complement of the locally constant functions. Then the bounds on the matrix coefficients show that there is  $\rho > 0$ , depending only on the spectral gap, so that:

$$|\langle s, \text{Pr}_0 \pi_* \chi, \text{Pr}_0 \pi_* \chi \rangle| \ll \varphi_0^{\text{wk}}(s)^\rho \quad (s \in G).$$

Recall that  $f$  was defined as the integral of the right translate of  $\chi$  by  $s_1^{-1}$  with  $s_1 \in B_2$ , which obviously give the same description of  $\pi_* f$  in terms of  $\pi_* \chi$ . Therefore, we have proved the claims (12.1) and (12.2) – at least with certain instances of  $\tilde{B}$  replaced by  $B_2$ . But it is easy to see that this is harmless because of the inclusion  $B_2 \subset \tilde{B}$ .  $\square$

### 13. EFFECTIVE CLOSING LEMMA.

Let us recall the closing lemma for hyperbolic flows. Let  $M$  be a compact manifold; let  $h_t : M \rightarrow M$  be a one-parameter flow of smooth diffeomorphisms with hyperbolicity transverse to the flow direction. Suppose that  $x \in M$  is so that the distance between  $h_T x$  and  $x$ , measured w.r.t. a fixed Riemannian metric on  $M$ , is sufficiently small. Then there exists  $y$  close to  $x$  and  $T'$  close to  $T$  so that  $h_{T'} y = y$ , i.e.  $y$  has a periodic orbit under  $\{h_t\}$ .

We shall need a method for producing periodic  $S$ -orbits, which is, in a certain sense, an analogue of this result. This is Proposition 13.1 (“Proposition C2” from the outline) – it asserts that, if  $x \in \Gamma \backslash G$  and if we are given a “sufficiently large” collection of elements  $s_i \in S$  so that  $xs_i$  are all mutually close to each other, then there exists  $x'$  near  $x$  so that  $x'S$  is closed.

**13.1. Proposition.** *Let  $H \subset S \subset G$ , with  $S$  connected. Let  $\delta \leq 1 \leq N$ . There exists some  $T_0 = T_0(\Gamma, G, H, N)$  with the following property:*

*Let  $T \geq T_0$  and let  $v = \text{vol } B_S(T)$ . Suppose that  $\{s_1, \dots, s_k\} \subset B_S(T)$  is  $\frac{1}{10}$ -separated, that  $k \geq v^{1-\delta}$ , and that there exists  $x \in X_{\text{cpt}}$  so that  $xs_i \stackrel{T^{-N}}{\sim} xs_j$ . Then there is  $x' \stackrel{T^{-N\uparrow}}{\sim} x$  so that  $x'S$  is a closed orbit of volume  $\leq T^{\delta\downarrow}$ .*

Recall that the  $\uparrow, \downarrow$  notation was defined in §3. In particular, the result is vacuous unless  $\delta$  is small enough.

**13.2. Preparations for the proof.** In the course of proof, we shall use the phrase  $T$  sufficiently large to mean that  $T \geq T_0(\Gamma, G, H, N)$ ; we are free to assume that  $T$  is sufficiently large. This allows us to replace any term of the form  $cT^{N\uparrow}$  by  $T^{N\uparrow}$ . Similarly, we shall say  $N$  is sufficiently large if  $N \geq N_0(G, H)$ . In view of the notation  $N_\uparrow$ , we are free to assume that  $N$  is sufficiently large.

Fix  $g_0 \in G$  so that  $\Gamma g_0 = x$ . Here  $g_0$  may be chosen in a compact subset of  $G$ , depending only on  $\Gamma, G$  (see discussion subsequent to Lemma 3.6.1.) There is  $\gamma_{ij} \in \Gamma$  so that  $\gamma_{ij} g_0 s_i \stackrel{T^{-N}}{\sim} g_0 s_j$  inside  $G$ . Thus, by (3.4),

$\gamma_{ij} \stackrel{T^{-N\uparrow}}{\sim} g_0 s_j s_i^{-1} g_0^{-1}$  for sufficiently large  $T$  and  $N$ . The assumption  $s_i \in B_S(T) \implies \|s_i\| \leq T$  together with (3.3) implies that  $\|\gamma_{ij}\| \leq c_1 T^2$  and  $\gamma_{jk} \gamma_{ij} \stackrel{T^{-N\uparrow}}{\sim} \gamma_{ik}$ ; which implies that  $\gamma_{jk} \gamma_{ij} = \gamma_{ik}$  if  $T$  and  $N$  are sufficiently large.

To sum up, in the setting of Proposition 13.1 we have produced a collection of elements  $\gamma_{ij} \in \Gamma$  so that:

$$(13.1) \quad \|\gamma_{ij}\| \leq c_1 T^2$$

$$(13.2) \quad \gamma_{jk} \gamma_{ij} = \gamma_{ik}.$$

$$(13.3) \quad \gamma_{ij} g_0 s_i \stackrel{T^{-N}}{\sim} g_0 s_j$$

Our proof proceeds as follows: First, we show we may slightly adjust  $g_0$  to a nearby  $g_1$  so that, in fact,  $\gamma_{ij} \in g_1 \tilde{S} g_1^{-1}$ . Here  $\tilde{S}$  is the normalizer of  $S$  as defined in §3.4. Next, we show that  $x' = \Gamma g_1$  has the property required:  $x'S$  is a closed orbit of small volume.

**13.3. Proof of Proposition 13.1.** We observe in advance that we will defer two results in the course of proof to the end of the section, to avoid interrupting the flow.

Let  $e_1, \dots, e_d$  be a basis for  $\mathfrak{s}$ , the Lie algebra of  $S$ . Let  $V := (\wedge^{\dim(\mathfrak{s})} \mathfrak{g})^{\otimes 2}$ , and set  $v_S = (e_1 \wedge \dots \wedge e_d)^{\otimes 2} \in V_{\mathbb{R}} := V \otimes_{\mathbb{Q}} \mathbb{R}$ . Then  $\mathbf{G}$  acts on the  $\mathbb{Q}$ -vector space  $V$ ,  $G$  acts on the  $\mathbb{R}$ -vector space  $V_{\mathbb{R}}$ , and the stabilizer of the vector  $v_S$  is precisely  $\tilde{S}$ ; the orbit  $G.v_S$  is a smooth submanifold of  $V_{\mathbb{R}}$ , as follows from general facts about orbits for algebraic groups.

There exists a  $G$ -equivariant projection map from an open neighbourhood of  $G.v_S$  to  $G.v_S$ . A precise statement and self-contained discussion of what we need in §13.5. See [34, Theorem 2.7] for much stronger and general statements.

For any finite subset  $F$  of the  $\gamma_{ij}$ s, let

$$(13.4) \quad X(F) = \{g \in G : F \subset g\tilde{S}g^{-1}\}.$$

(We work with  $\tilde{S}$  instead of  $S$  only to make the following argument as explicit as possible.) There is an integer  $\kappa_{25}$ , depending only on  $G$  and  $H$  so that:

$$(13.5) \quad F \subset \{\gamma_{ij}\}, \quad |F| \leq \kappa_{25}, \quad X(F) = X(\{\gamma_{ij}\}).$$

This is an example of an “effective” Noetherian argument. Start with  $F = \emptyset$  and adjoin one  $\gamma_{ij}$  at a time to  $F$ . The resulting sets  $X(F)$  are the real points of a descending sequence of algebraic varieties, which must terminate; (13.5) gives an explicit estimate for how far one must go. We give a precise argument in §13.6. Take such an  $F = \{\delta_1, \dots, \delta_m\}$ , where  $m \leq \kappa_{25}$ .

The Euclidean norm on  $\mathfrak{g}$  induces one on  $V_{\mathbb{R}}$  and  $V_{\mathbb{R}}^m$ . Consider the map

$$V_{\mathbb{R}} \xrightarrow{A} V_{\mathbb{R}}^m$$

where  $A = \oplus_{i=1}^m (\delta_i - 1)$ . The element  $g_0 v_S$  almost belongs to the kernel of  $A$ : in fact  $\|A.g_0 v_S\| \ll T^{-N\uparrow}$ , as follows from (13.3) and the fact  $\|s_i\| \leq T$ .

Moreover, with respect to a fixed basis for the  $\mathbb{Q}$ -vector space  $V$ , all the entries of  $A$  are rational numbers of numerator and denominator  $\leq T^*$ . This follows from (13.1), and the fact that, since  $\Gamma$  is arithmetic, the denominators of matrix entries of  $A$  are bounded below.

It follows that (see Lemma 13.4.1 at the end of the present section for an explication) there is a nearby vector in the kernel of  $A$  which belongs to the kernel, i.e. an element  $v'_S \in V_{\mathbb{R}}$  such that

$$(13.6) \quad \delta_i v'_S = v'_S, \quad 1 \leq i \leq m \text{ and } \|v'_S - g_0 v_S\| \ll T^{-N_{\uparrow}}.$$

We claim that there exists  $g_1$ , satisfying  $g_1 \stackrel{T^{-N_{\uparrow}}}{\sim} g_0$  for large enough  $T$  and  $N$ , so that  $v'_S := g_1 v_S$  also satisfies (13.6). This claim will imply that  $g_1 \in X(F) \neq \emptyset$  (see (13.5)) and therefore that  $\gamma_{ij} \in g_1 \tilde{S} g_1^{-1}$ .

We have noted that  $g_0$  may be chosen to belong to a compact subset of  $G$  depending only on  $H$  and  $\Gamma$ . Therefore,  $v_S$  and so also  $g_0^{-1} v'_S$  belongs to a fixed compact subset of  $V_{\mathbb{R}}$  depending only on  $\Gamma$ ,  $G$ ,  $H$ , and  $S$ . The equivariant projection map  $\Pi$  mentioned above restricted to this subset is Lipschitz, with a constant depending only on  $\Gamma$ ,  $G$ ,  $H$ , and  $S$ . Therefore – replacing  $v'_S$  by  $g_0 \Pi(g_0^{-1} v'_S)$  – we may assume that  $v'_S \in G.v_S$  without changing the fact that  $\|v'_S - g_0 v_S\| \ll T^{-N_{\uparrow}}$ . Next, the map  $g \mapsto g.v_S$  is a submersive map from  $G$  to  $G.v_S$ , in a neighbourhood of  $g_0$ . We may thereby find  $g_1$  near  $g_0$  so that  $v'_S = g_1 v_S$ . The claim after (13.6) follows, at least for  $T$  sufficiently large.

Having modified  $g_0$  to a nearby  $g_1$ , we now proceed to show that  $x' = \Gamma g_1$  has the properties required by the Proposition. Precisely, we shall show that:

$$(13.7) \quad \Lambda := g_1^{-1} \Gamma g_1 \cap S \text{ is a lattice within } S \text{ of covolume } \ll T^{\delta_{\downarrow}}.$$

This will establish the proposition.

By construction,  $g_1 \stackrel{T^{-N_{\uparrow}}}{\sim} g_0$  and  $\gamma_{ij} \in g_1 \tilde{S} g_1^{-1}$ . Set  $\gamma'_{ij} = g_1^{-1} \gamma_{ij} g_1$ ,  $s'_i = \gamma'_{1i} s_1$ . Then  $\gamma'_{ij}, s'_i \in \tilde{S}$  and  $s'_i = \gamma'_{ji} s'_j$  by (13.2) for all pairs  $i, j$ . Also,  $s'_i \stackrel{T^{-N_{\uparrow}}}{\sim} s_i$  for sufficiently large  $T$ , because  $s_i \stackrel{T^{-N_{\uparrow}}}{\sim} g_0^{-1} \gamma_{1i} g_0 s_1$  (see (13.3)). Consequently, for  $T$  sufficiently large, the  $s'_i$  are  $\frac{1}{20}$ -separated and belong to  $B_S(c_2 T)$ .<sup>27</sup>

We are going to apply the upper bound (part 2) of Proposition 12.2 with  $\Lambda$  as in (13.7),  $B := B_S(c_2 T)$  and with  $\{s'_i\}$  as the separated subset of  $B \cap \Lambda s'_1$ . By the remark after Proposition 12.2, the proposition also yields bounds for the cardinality of any  $\frac{1}{20}$ -separated subset, such as  $\{s'_i\}$ ; the upper bound is weaker by a constant that depends only on  $H, G$ . Notations as in that proposition, the ball “ $\tilde{B}$ ” is contained in  $B_S(c_3 T)$ , for  $T$  sufficiently large, as we see by applying property (3) of §3.5.

<sup>27</sup>Indeed, the fact that  $s'_i \stackrel{T^{-N_{\uparrow}}}{\sim} s_i$  forces  $s'_i = s_i \omega_i$ , where  $\omega_i \stackrel{T^{-N_{\uparrow}}}{\sim} 1$ . For  $T$  sufficiently large, this forces  $\omega_i \in S$ . Therefore, by property (3) of §3.5 and the fact  $s_i \in B_S(T)$ , we conclude that  $s' \in B_S(c_2 T)$ .

In order to apply that Proposition, we apply Proposition 6.1 with  $x = \Gamma g_1$  to show that the action of  $S$  on the orthogonal complement of the locally constant functions in  $L^2(\Lambda \backslash S)$  has a spectral gap (depending only on  $G, H$ ). Let  $\rho$  be as in the statement of Proposition 12.2; it depends only on this spectral gap, and, therefore, only on  $H, G$ .

We note that, by results established in §6.4, we have

$$(13.8) \quad \frac{1}{\text{vol } B_S(c_3 T)^2} \int_{B_S(c_3 T)} (\varphi_0^{\text{wk}}(g_1 g_2^{-1}))^\rho dg_1 dg_2 \ll T^{-\zeta_S \rho} \ll v^{-\frac{\zeta_S \rho}{2A_S}}$$

for suitable  $\zeta_S, A_S > 0$  depending on  $S$ .

Apply Proposition 12.2, (2). We conclude that:

$$v^{1-\delta} \leq |\{s_i\}| \ll \frac{v}{\sqrt{\text{vol}(\Lambda \backslash S)}} + vv^{-\zeta_S \rho / 4A_S}$$

In other terms,  $\text{vol}(\Lambda \backslash S)^{-1/2} + v^{-\frac{\zeta_S \rho}{4A_S}} \gg v^{-\delta}$ . In particular, if  $\delta < \zeta_S \rho / 4A_S$  and  $T$  is sufficiently large, we see that  $\text{vol}(\Lambda \backslash S) \ll v^{2\delta}$ . In particular,  $\Lambda$  must be a lattice in  $S$ : our conventions dictated that its covolume is  $\infty$  if this is not the case.

This concludes the proof of (13.7), and therefore the Proposition.  $\square$

We now establish some (simple) results that were used in the above proof.

**13.4. A vector almost in the kernel of a rational matrix, is near a vector in the kernel.** In the following statement,  $\|\cdot\|$  refers to the standard Euclidean norm on  $\mathbb{R}^n$  and  $\mathbb{R}^m$ .

**13.4.1. Lemma.** *Let  $A \in M_{n \times m}(\mathbb{Z})$  be an  $n \times m$  integer matrix, all of whose entries are  $\leq E$  in absolute value. Suppose  $v \in \mathbb{R}^m$  with  $\|Av\| \leq \delta$ . Then there exists  $v_0 \in \ker(A)$  with  $\|v - v_0\| \leq \delta(nm)^{n/2} E^n$ .*

*Proof.* There exists an orthonormal basis  $v_1, \dots, v_m$  for  $\mathbb{R}^m$  with the property that  $Av_i$  are orthogonal in  $\mathbb{R}^n$  and so that the lengths  $\sigma_i := \|Av_i\|$  are decreasing. The  $\sigma_i$  are the singular values and  $\sigma_i^2$  are the eigenvalues of  $AA^t$  (see §7.2).

The matrix  $AA^t$  is an  $n \times n$  matrix, all of whose entries are integers  $\leq mE^2$  in absolute value. The absolute value of any eigenvalue is thereby bounded by  $nmE^2$ . Moreover, the eigenvalues are algebraic integers; in particular, the product of all their algebraic conjugates is a nonzero integer. It follows from this that the absolute value of any nonzero eigenvalue of  $AA^t$  is bounded below by  $(nmE^2)^{-n}$ . Therefore, each nonzero  $|\sigma_i| \geq (nmE^2)^{-n/2}$ .

Now take  $v \in \mathbb{R}^m$  so that  $\|Av\| \leq \delta$ . The projection  $v'$  of  $v$  onto the span of all  $v_i$  with  $Av_i \neq 0$  therefore has length  $\leq \delta(nm)^{n/2} E^n$ . Set  $v_0 = v - v'$ .  $\square$

**13.5. Existence of an equivariant projection.** Notations as in the proof. We shall show the existence of an equivariant projection from a fixed small neighbourhood of  $v_S$ , onto  $G.v_S$ . This suffices for the argument in the text, although one can do much better, see [34].

The map  $g \mapsto g.v_S$  induces a map on tangent spaces and in particular a map  $\mathfrak{g} \rightarrow V_{\mathbb{R}}$ . Since  $S$  is semisimple, we may choose an  $S$ -invariant complement  $W$  to the image of  $\mathfrak{g}$ . Let  $G \times_S W$  be the quotient of pairs  $(g, w)$  by the equivalence  $(gs, s^{-1}w) \sim (g, w)$ ; it is a vector bundle over  $G/S$  and so a manifold. Consider the map  $G \times_S W \rightarrow V_{\mathbb{R}}$ , given by  $(g, w) \mapsto g.(v_S + w)$ . The differential of this map is an isomorphism at  $(g, w) = (1, 0)$  and therefore, by the implicit function theorem, it defines a diffeomorphism from a neighbourhood  $\mathcal{N}_1 \in G \times_S W$  of  $(1, 0)$ , to a neighbourhood  $\mathcal{N}_2$  of  $v_S$ . In this neighbourhood there exists a projection that is, at least locally, equivariant, which is given by

$$(13.9) \quad \Pi : g(v_S + w) \in \mathcal{N}_2 \mapsto gv_S.$$

This crude construction already suffices for our simple application. Indeed, it implies the following weak Lemma, which is easily seen to imply our application:

**13.5.1. Lemma.** *There exists a constant  $\kappa_{26} > 0$  with the following property. Let  $v \in \mathcal{N}_2$  and  $\delta \in \Gamma$ . If  $\delta v = v$ , and  $\|v - v_S\| \leq \|\delta\|^{-\kappa_{26}}$ , then also  $\delta\Pi(v) = \Pi(v)$ .*

*Proof.* Choose  $(g, w) \in \mathcal{N}_1$  so that  $g(v_S + w) = v$ . Observe that  $\|gv_S - v_S\|, \|gw\| \ll \|v_S - v\|$ .

Then  $\Pi(v) = gv_S$ . Since  $(\delta v - v) = (\delta gv_S - v_S) + (v_S - gv_S) + (\delta gw - gw)$ , we get

$$(13.10) \quad \|\delta gv_S - v_S\| \leq \|gv_S - v_S\| + \|\delta(gw) - (gw)\| \ll \|\delta\|^* \|v_S - v\|.$$

From this, we see that  $(\delta g, w) \in \mathcal{N}_1$  if  $\kappa_{26}$  was chosen big enough. (Indeed,  $(\delta g, w) = (\delta gs, s^{-1}w)$  in  $G \times_S W$ . The assumption  $\|v - v_S\| \leq \|\delta\|^{-\kappa_{26}}$  and (13.10) ensure that  $(\delta gs, s^{-1}w)$  is close to  $(1, 0)$  for a suitable choice of  $s$ .)

However,  $(\delta g, w)$  and  $(g, w)$  both map to  $\delta v = v$  under the diffeomorphism from  $\mathcal{N}_1$  to  $\mathcal{N}_2$ . Therefore,  $\delta gS = gS$ , i.e.  $\delta\Pi(v) = \Pi(v)$ .  $\square$

**13.6. Effective noetherian arguments.** Set as before  $X(F) = \{g \in G : F \subset g\tilde{S}g^{-1}\}$ . We shall show that, given any set  $F_1$ , there exists a subset  $F \subset F_1$  of size bounded only in terms of  $G, S$ , so that  $X(F_1) = X(F)$ .

Indeed, notations as before, we have:

$$X(\{\alpha\}) = \{g \in G : \alpha gv_S = gv_S\} = \{g \in G : gv_S \in \text{Fix}(\alpha)\}$$

where  $\text{Fix}(\alpha)$  is the fixed locus of  $\alpha$  acting on  $V$ .

Given a vector space of dimension  $d$ , and a collection of linear subspaces, any intersection of these subspaces can in fact be expressed as the intersection of at most  $d$  of them. In particular, there exists  $F_1 \subset F$  of size  $\leq \dim V$  so that:

$$\bigcap_{\alpha \in F_1} \text{Fix}(\alpha) = \bigcap_{\alpha \in F} \text{Fix}(\alpha).$$

This implies that  $X(F_1) = X(F)$ , as desired.  $\square$

## 14. A COROLLARY OF THE CLOSING LEMMA.

14.1. In the previous section, we presented a version of a closing lemma for actions of semisimple groups. In the present section, we shall apply it to the setting needed in this paper, proving the following (“Proposition C” from §4).  $\mu$  is as in Theorem 1.3.

**14.2. Proposition.** *Given a connected intermediate subgroup  $H \subseteq S \subseteq G$ ,  $\zeta \in (0, 1)$ , and  $d_S \geq 1$ , there exists  $\xi$  and  $d'$  depending only on  $d_S, \zeta, G, H$ , and  $\epsilon_0 \ll_{d_S, \zeta} 1$ , so that:*

*Suppose for some  $\epsilon \leq \epsilon_0$  that*

- (1)  *$\mu$  is  $\epsilon$ -almost invariant under  $S$ , with respect to the Sobolev norm  $\mathcal{S}_{d_S}$ , and that*
- (2)  *$\mu(xS) = 0$  for all closed  $S$ -orbits of volume  $\leq \epsilon^{-\zeta}$ .*

*Then there exists  $x_1, x_2$  so that  $x_2 = x_1 \exp(r)$ ,  $r \in \mathfrak{r}$ ,  $\|r\| \leq \epsilon^\xi$ , and  $x_1, x_2$  are both  $[\|r\|^{-\kappa_8}, \|r_1\|^{-\kappa_9}]$ -generic w.r.t.  $\mathcal{S}_{d'}$ .*

We strongly suggest the reader glance at the description of this argument, provided in §4.5, prior to reading what follows. Also recall that  $r_1$  stands for the non-fixed components of  $r$  in the splitting defined in (3.2).

**14.3. Beginning of the proof.** In this proof,  $\epsilon$  sufficiently small will mean that  $\epsilon$  is bounded above by a constant  $c_1(d_S, \zeta)$ . In view of the formulation of the Proposition, we are free to assume  $\epsilon$  is sufficiently small.

Let  $\beta$  as in Proposition A (9.3) with  $d = d_S$ . Let  $A_S$  be as in (6.12).

Let  $N_\uparrow, \delta_\downarrow$  be the functions defined in Proposition 13.1. Take  $N$  so large that  $N_\uparrow \geq 2A_S\kappa_{12}$  where  $\kappa_{12}$  is as in Proposition C1 (11.1). Choose  $\delta$  so small that  $\delta_\downarrow \leq A_S/2$ . Let  $q$  be chosen so small that:

- $q \leq \beta$ ;
- $2qN\kappa_9 < \beta$ .
- $2qA_S \leq \zeta$ .
- $qd_S\kappa_{22} < \frac{1}{2}$ .

Finally, we put  $\xi = \frac{\delta q A_S}{5 \dim(G)}$ . These choices of  $N, \delta, q$  depend only on  $G, H, \zeta, d_S$ . We set  $T = \epsilon^{-q}$  and  $v = \text{vol}(B_S(T))$ . Our constraints imply that, for  $\epsilon$  sufficiently small, we have

$$(14.1) \quad T^{N_\uparrow} \geq v^{\kappa_{12}}, \quad T^{\delta_\downarrow} \leq v, \quad v^{-\frac{\delta}{2 \dim(G)}} \leq \epsilon^{2\xi}, \quad T \leq \epsilon^{-\beta}, \quad v \leq \epsilon^{-\zeta}.$$

The reader should not pay too much attention to the mass of constants above: simply,  $N$  and  $\delta$  are *fixed*;  $\epsilon$  is *very small*, and  $v^{-1}$  is *small*. Set

(14.2)

$$E_1 = \{x : \text{there is } x' \stackrel{v^{-\kappa_{12}}}{\sim} x \text{ such that } x'S \text{ is closed of volume } \leq v\}.$$

Under our assumption, Proposition C1 (11.1) furnishes an upper bound for the  $\mu$ -measure of  $E_1$ . In fact,

$$(14.3) \quad \mu(E_1) \leq 1/2 \text{ for } v \in [V_0(\Gamma, G, H), \epsilon^{-\zeta}].$$

Note that the condition on  $v$  will be satisfied if  $\epsilon$  is sufficiently small.

Apply Proposition A (9.3) with  $d = d_S$ ,  $R = T$ . It produces points generic with respect to a new Sobolev norm  $\mathcal{S}_{d'}$ ; generic is always understood in that sense, in what follows. More precisely, Proposition 9.3 asserts that the fraction of pairs  $(x \in X, s \in B_S(T))$  so that  $x.s$  is not  $[T_0, \epsilon^{-\beta}]$ -generic is  $\ll_{d_S} T_0^{-1}$ . We choose  $T_0$  large enough so that this fraction of points is  $< \frac{1}{10^{10}}$ . This choice of  $T_0$  depends only on  $\Gamma, G, H, d_S$ .

Next, we observe that:

$$(14.4) \quad \text{The fraction of } (x \in X, s \in B_S(T)) \text{ for which } x.s \notin X_{\text{cpct}} \text{ is } \leq \frac{1}{10^{10}}$$

if  $\epsilon$  is sufficiently small. Here “fraction” is measured w.r.t.  $\mu \times \text{vol}_S$ . Indeed, recalling that  $X_{\text{cpct}} = \mathfrak{S}(R_0)$ , take a smooth function  $F$  so that:

$$1_{X - \mathfrak{S}(R_0/2)} \geq F \geq 1_{X - \mathfrak{S}(R_0)}$$

Invoking Lemma 8.2.2 and the almost invariance,

$$\left| \int F(xs) d\mu(x) - \int F d\mu \right| \ll \epsilon T^{d_S \kappa_{22}} \mathcal{S}_{d_S}(F) \text{ whenever } s \in B_S(T).$$

Thus

$$\frac{\int_{s \in B_S(T), x \in X} F(xs)}{\text{vol } B_S(T)} - \mu(X - \mathfrak{S}(R_0/2)) \ll \epsilon^{1 - q d_S \kappa_{22}} \mathcal{S}_{d_S}(F).$$

$F$  may be fixed in a fashion depending only on  $\Gamma, G$ . Moreover,  $1 - q d_S \kappa_{22} > 1/2$ . Therefore, if  $\epsilon$  sufficiently small, the observation (14.4) follows.

For each  $x \in X$ , let

$$f(x) = \frac{1}{\text{vol}(B_S(T))} \text{vol}(\{s \in B_S(T) : xs \in X_{\text{cpct}}, xs \text{ is } [T_0, \epsilon^{-\beta}]\text{-generic}\})$$

The function  $f$  takes values in  $[0, 1]$ . In view of our remarks above,  $\int (1 - f(x)) d\mu \leq \frac{2}{10^{10}}$ . Thus, the set  $E_2 = \{x \in X : f(x) < 1 - 10^{-6}\}$  satisfies  $\mu(E_2) < \frac{1}{10}$ . Let  $X_{\text{good}} = X_{\text{cpct}} - E_1 - E_2$ .

The set  $X_{\text{good}}$  is a  $\mu$ -nonempty set and has the following properties:

- (1)  $X_{\text{good}} \subset X_{\text{cpct}}$ .
- (2) For any  $x \in X_{\text{good}}$ , the set

$$(14.5) \quad \mathcal{B}_x = \{s \in B_S(T) : xs \text{ is } [T_0, \epsilon^{-\beta}]\text{-generic and belongs to } X_{\text{cpct}}\}$$

has measure larger than  $(1 - 10^{-6}) \text{vol } B_S(T)$ .

- (3) For any  $x \in X_{\text{good}}$ , there does not exist  $x' \sim^{v^{-\kappa_{12}}} x$  such that  $x'S$  is closed of volume  $\leq v$ .

Let  $x \in X_{\text{good}}, \mathcal{B}_x$  as in (14.5).

**Lemma.** *There exists  $b_1, b_2 \in \mathcal{B}_x$  so that  $xb_1 = xb_2 \exp(r)$  for some  $r \in \mathfrak{r}$ , where  $v^{-\frac{\delta}{2 \dim(G)}} \gg \|r\| \gg T^{-N}$ . Moreover,  $\|r_1\| \gg \|r\|$ .*

The Proposition follows from this, as we now explicate:

By choice,  $v^{-\frac{\delta}{2 \dim(G)}} \leq \epsilon^{2\xi}$ . Then  $\|r\| \leq \epsilon^\xi$ , for  $\epsilon$  sufficiently small. Let us verify that

$$[\|r\|^{-\kappa_8}, \|r_1\|^{-\kappa_9}] \subseteq [T_0, \epsilon^{-\beta}].$$

Since  $T_0 \ll_{d_S} 1$ , the inequality  $\|r\|^{-\kappa_8} \geq T_0$  holds if  $\epsilon$  is sufficiently small. The other inequality  $\|r_1\|^{-\kappa_9} \leq \epsilon^{-\beta}$  follows for sufficiently small  $\epsilon$  because  $\|r_1\|^{-\kappa_9} \ll \epsilon^{-qN\kappa_9}$ , and  $qN\kappa_9 \leq \beta/2$ .

**14.4. Proof of the lemma.** Let us recollect the situation. We are given  $x \in X_{\text{cpct}}$  and a subset  $\mathcal{B}_x \subset B_S(T)$  satisfying  $\frac{\text{vol}(\mathcal{B}_x)}{\text{vol } B_S(T)} \geq 1 - 10^{-6}$ , and we know that there does not exist  $x' \stackrel{v^{-\kappa_{12}}}{\sim} x$  such that  $x'S$  is closed of volume  $\leq v$ .

We are free to prove the Lemma for  $T$  sufficiently large: here and in the course of the proof, the phrase “for  $T$  sufficiently large” to mean “for  $T$  larger than a constant that may depend on  $\delta, N, \Gamma, G$ , and  $H$ .” Indeed, we can guarantee  $T$  sufficiently large by taking  $\epsilon$  sufficiently small.

Take *sufficiently small* symmetric neighbourhoods  $\Omega'$ , and  $\Omega$  of the identity in  $S$ , resp.  $\Omega_{\mathfrak{r}}$  of  $0 \in \mathfrak{r}$  such that:

$$(14.6) \quad \Omega' \subset \Omega \subset \{s \in S : d(s, e) \leq 1/4\} \subset S.$$

The precise notion of *sufficiently small* will be specified in the course of the argument; however, it will depend only on  $\Gamma, G, H$  and therefore constants that depend on  $\Omega', \Omega, \Omega_{\mathfrak{r}}$  will be absorbed into  $\ll$  notation.

Set

$$(14.7) \quad \mathcal{B} = \{s \in \mathcal{B}_x : \text{vol}(s\Omega \cap \mathcal{B}_x) \geq 0.99 \text{vol}(\Omega)\}$$

We shall verify that, for  $T$  sufficiently large, we have also

$$(14.8) \quad \text{vol}(\mathcal{B}) \gg \text{vol}(B_S(T)) (= v).$$

If we set  $f(s) = \frac{\text{vol}(s\Omega \cap \mathcal{B}_x)}{\text{vol}(\Omega)}$ , then  $\int_{\mathcal{B}_x} f(s) ds = (\text{vol}(\Omega))^{-1} \int_{\Omega} \text{vol}(\mathcal{B}_x s \cap \mathcal{B}_x) ds$ , which exceeds

$$\text{vol}(\Omega)^{-1} \int_{\Omega} (\text{vol}(B_S(T)s \cap B_S(T)) - 10^{-5} \text{vol}(B_S(T))) ds.$$

By property (3) of §3.5, there is a constant  $c < 1$  so that  $B_S(cT) \subset B_S(T)s \cap B_S(T)$  for any  $s \in S$  with  $d(s, e) \leq 1/4$ . In particular, using (6.12), property (3b) of §3.5, and choosing  $\Omega$  sufficiently small, we can arrange that

$$(14.9) \quad \text{vol}((B_S(T)s \cap B_S(T))) \geq (1 - 10^{-5}) \text{vol } B_S(T)$$

for any  $s \in \Omega$ , at least for  $T \gg 1$ . Thus,  $\int_{\mathcal{B}_x} f(s) ds > (1 - 10^{-4}) \text{vol } B_S(T)$ ; in particular, the set of  $s \in B_S(T)$  for which  $f(s) > 0.99$  has volume at least  $\frac{1}{2} \text{vol } B_S(T)$ .

For  $y \in X$ , let

$$N(y) := \text{vol}(\{s \in \mathcal{B} : xs\Omega \cap B(y, T^{-N}/2) \neq \emptyset\})$$

Loosely speaking,  $N(y)$  measures the number of “times” in  $\mathcal{B}$  for which the corresponding point in  $x\mathcal{B}$  comes within  $T^{-N}$  of  $y$ . We consider two separate cases:

- Case I : For every  $y \in X$ ,  $N(y) \leq v^{1-\delta/2}$ .  
 Case II : There is  $y \in X$  with  $N(y) \geq v^{1-\delta/2}$ .

Let us show that Case II cannot occur by using that  $x$  is not close to a closed orbit of small volume. Suppose that Case II occurred; let  $y_0 \in X$  be so that  $N(y_0) \geq v^{1-\delta/2}$ . Choose a maximal 1-separated subset  $\{s_i\}_{1 \leq i \leq I}$  of

$$\{s \in \mathcal{B} : xs\Omega \cap B(y_0, T^{-N}/2) \neq \emptyset\}$$

Then  $I \gg v^{1-\delta/2}$ . In explicit terms, there are  $\omega_i \in \Omega$  ( $i \in I$ ) so that

$$xs_i\omega_i \stackrel{T^{-N}}{\sim} xs_j\omega_j \quad (i, j \in I).$$

In particular, the elements  $s'_i = s_i.\omega_i$  are  $1/2$ -separated<sup>28</sup> and belong to  $B_S(c_2T)$ . In view of our assumptions on  $\delta, N$ , Proposition C2 (13.1) – applicable so long as  $\epsilon$  is sufficiently small – would show there is  $x' \stackrel{v^{-\kappa_{12}}}{\sim} x$  so that  $x'S$  is a closed orbit of volume  $\leq v$ ; but that contradicts our assumption on the point  $x$ .

Therefore, we are in Case I. It will be convenient to pass to a local coordinate system. Clearly, there is  $x' \in X_{\text{cpct}}$  so that:

$$\text{vol}\{s \in \mathcal{B} : xs \in x' \exp(\Omega_{\mathfrak{r}})\Omega'\} \gg v.$$

Indeed, we may cover  $X_{\text{cpct}}$  by finitely many neighborhoods of the form  $x' \exp(\Omega_{\mathfrak{r}})\Omega'$ ; and, by assumption,  $x\mathcal{B} \subset X_{\text{cpct}}$ .

For each  $s \in \mathcal{B}$  so that  $xs$  belongs to our chosen neighbourhood, we may write:

$$xs = x' \cdot \exp(r_s) \cdot \sigma_s, \quad r_s \in \Omega_{\mathfrak{r}}, \sigma_s \in \Omega'.$$

Let  $B$  be any metric ball of radius  $T^{-N}/4$  in the Euclidean space  $\mathfrak{r}$ , so that  $B \cap \Omega_{\mathfrak{r}} \neq \emptyset$ . We claim that the preimage of  $B$ , under the map  $s \mapsto r_s$ , has measure  $\leq v^{1-\delta/2}$ . Were this false, there exists a subset  $\mathcal{B}' \subset \mathcal{B}$  of volume  $> v^{1-\delta/2}$ , so that  $x\mathcal{B}' \subseteq x' \exp(B)\Omega'$ . If  $\Omega_{\mathfrak{r}}$  is sufficiently small,  $x' \exp(B) \subset B(x'', T^{-N}/2)$ , for a suitable  $x'' \in x' \exp(B)$ .

In particular,  $N(x'') > v^{1-\delta/2}$ , in contradiction to the assumption in Case I.

Cover  $\Omega_{\mathfrak{r}}$  by  $T^{-N}/4$ -balls so that each one overlaps with  $O(1)$  others. The previous paragraph shows that  $\gg v^{\delta/2}$  of these balls contain a point of the form  $r_s$ , for some  $s \in \mathcal{B}$ . Thus there is  $s_1, s_2$  so that  $T^{-N} \ll \|r_{s_1} - r_{s_2}\| \ll v^{-\delta/(2 \dim(\mathfrak{r}))}$ . Thus,

(14.10)

$$xs_1 = x' \exp(r_{s_1})\sigma_1, \quad x's_2 = x \exp(r_{s_2})\sigma_2, \quad T^{-N} \ll \|r_{s_1} - r_{s_2}\| \ll v^{-\delta/(2 \dim(\mathfrak{r}))}$$

and  $\sigma_1, \sigma_2 \in \Omega' \subset S$ .

<sup>28</sup>Indeed,  $d(s_i, s'_i) \leq 1/4$  by the left-invariance of the metric, and  $d(s_i, s_j) \geq 1$  by choice.

Using (14.10), we shall now perturb  $s_1$  and  $s_2$  along  $S$  slightly to find  $b_1, b_2 \in \mathcal{B}_x$  as in the first claim of the lemma. For  $i \in \{1, 2\}$ , let  $W_i = \{s \in \Omega : s_i s \in \mathcal{B}_x\}$ . Then  $\text{vol}(W_i) > 0.9 \text{vol}(\Omega)$  by (14.7) and the surrounding discussion. Our (intuitively obvious) conclusion follows from the following

**Sublemma.** *Suppose we are given  $\sigma_i \in \Omega'$ ,  $r'_i \in \Omega_{\mathfrak{r}}$ , and subsets  $W_i \subset \Omega$  of measure  $\text{vol}(W_i) > 0.9 \text{vol}(\Omega)$  for  $i = 1, 2$ . Then there exists  $w_i \in W_i$  for  $i = 1, 2$  so that*

$$\exp(r'_1)\sigma_1 w_1 = \exp(r'_2)\sigma_2 w_2 \exp(r), \quad r \in \mathfrak{r}, \|r\| \asymp \|r'_1 - r'_2\|, \|r_1\| \gg \|r\|.$$

This statement easily implies the Lemma: First of all,  $b_i = s_i w_i \in \mathcal{B}_x$ . Therefore,  $x_i := x b_i$  satisfy the constraints of the Lemma (i.e.  $x_2 = x_1 \exp(r)$ , where  $v^{-\frac{\delta}{2 \dim(G)}} \gg \|r\| \gg T^{-N}$ .)

**14.5. Proof of the sublemma.** Define functions  $\phi_S \times \phi_{\mathfrak{r}}(s) : \Omega \rightarrow S \times \mathfrak{r}$  according to:

$$(14.11) \quad (\exp(r'_2)\sigma_2)^{-1} \exp(r'_1)\sigma_1 s = \phi_S(s) \exp(\phi_{\mathfrak{r}}(s)), \quad \phi_S(s) \in S, \phi_{\mathfrak{r}}(s) \in \mathfrak{r}.$$

Strictly speaking, the maps  $\phi_S, \phi_{\mathfrak{r}}$  are functions of  $r'_1, r'_2, \sigma_1, \sigma_2$ , and  $s$ ; but we shall suppress the dependence on the first four variables, which will be fixed throughout the proof.

If  $\Omega, \Omega_{\mathfrak{r}}$ , and  $\Omega'$  are sufficiently small, the expression on the left of (14.11) is “sufficiently close” to the origin. This implies that, if we take the three sets  $\Omega, \Omega', \Omega_{\mathfrak{r}}$  sufficiently small,  $\phi_S$  and  $\phi_{\mathfrak{r}}$  will be defined and smooth. In that statement, the notion of “sufficiently small” depends only on  $G, H$ .

We require that  $\Omega, \Omega'$ , and  $\Omega_{\mathfrak{r}}$  are so small that:

- $\phi_S$  is injective,
- The preimage  $\Omega'' := \phi_S^{-1}(\Omega) \subset \Omega$  of  $\Omega$  has volume  $\text{vol}(\Omega'') > 0.9 \text{vol}(\Omega)$ .
- The Jacobian of  $\phi_S$  is everywhere on  $\Omega$  between 0.9 and 1.1, i.e. the map  $\phi_S$  almost preserves volume.

This is possible since the dependence of  $\phi_S$  on  $s$  and the parameters  $r'_1, r'_2, \sigma_1, \sigma_2$  is smooth and since  $\phi_S(s) = s$  if the latter parameters are trivial. Clearly these restrictions only depend on  $S$ .

These conditions imply  $\text{vol}(\phi_S(W_1 \cap \Omega'')) > 0.5 \text{vol}(\Omega)$ . Therefore,  $W_2 \cap \phi_S(W_1 \cap \Omega'')$  must be nonempty. Take  $w_1 \in W_1$  with  $\phi_S(w_1) = w_2 \in W_2$ . We may also assume that  $w_1$  is, in a weak sense, a point of density for this set:  $\phi_S^{-1}(W_2) \cap W_1$  intersects  $w_1 \Omega'$  in a set of volume bounded below.

Then:

$$\exp(r'_1)\sigma_1 w_1 = \exp(r'_2)\sigma_2 w_2 \exp(\phi_{\mathfrak{r}}(w_1)).$$

by (14.11).

We claim  $\|\phi_{\mathfrak{r}}(w_1)\| \asymp \|r'_1 - r'_2\|$ . To see that, rearrange (14.11) as follows:

$$\begin{aligned} \exp(r'_2)^{-1} \exp(r'_1) &= \\ \sigma_2 \phi_S(w_1) \exp(\phi_{\mathfrak{r}}(w_1)) (\sigma_1 w_1)^{-1} &\in S \exp(\text{Ad}(\sigma_1 w_1) \phi_{\mathfrak{r}}(w_1)) \end{aligned}$$

It will therefore suffice to prove that, if we express  $\exp(r'_2)^{-1} \exp(r'_1)$  in the form  $s \exp(X)$  ( $s \in S, X \in \mathfrak{r}$ ), we have the majorization  $\|X\| \gg \|r'_1 - r'_2\|$  for sufficiently small  $\Omega_{\mathfrak{r}}$ . This has already been seen: (10.1).

It remains to check that  $\|\phi_{\mathfrak{r}}(w_1)_1\| \gg \|\phi_{\mathfrak{r}}(w_1)\|$ . We will do this after possibly modifying  $w_1$ , using the “point of density” assumption. For  $\sigma \in S$  near to the identity, we have  $\phi_{\mathfrak{r}}(w_1 \sigma) = \text{Ad}(\sigma^{-1}) \phi_{\mathfrak{r}}(w_1)$ ; replacing  $w_1$  by a nearby  $w_1 \sigma \in \phi_S^{-1}(W_2) \cap W_1$ , and applying a suitable variant of (10.5), we conclude.

This concludes the proof of the Sublemma.  $\square$

# 15. AN ALMOST INVARIANT MEASURE ON A CLOSED ORBIT IS CLOSE TO THE INVARIANT MEASURE.

We shall prove (“Proposition F” from §4)

**15.1. Proposition.** *Let  $x_0 \in X$  be so that  $x_0 S$  is a closed orbit of volume  $V$ , for some connected  $S \supset H$ . Suppose  $\mu$  is a probability measure on  $x_0 S$  that is  $\epsilon$ -invariant under  $S$  w.r.t. a Sobolev norm  $\mathcal{S}_d$ . Let  $\nu$  be the  $S$ -invariant probability measure on  $x_0 S$ .*

*Then there are  $\kappa_{14}$  and  $\kappa_{15}$  so that*

$$(15.1) \quad |\mu(f) - \nu(f)| \ll_d V^{\kappa_{14}} \epsilon^{\kappa_{15}/d} \mathcal{S}_d(f) \text{ for } f \in C_c^\infty(X).$$

*In particular, there are constants  $\kappa_{16}, \kappa_{17} > 0$  such that if  $V \leq \epsilon^{-\kappa_{16}/d}$  then  $\mu$  and  $\nu$  are  $\epsilon^{\kappa_{17}/d}$ -close:*

$$|\mu(f) - \nu(f)| \ll_d \epsilon^{\kappa_{17}/d} \mathcal{S}_d(f) \text{ for } f \in C_c^\infty(X).$$

One cannot – at least naively – dispense with the occurrence of  $V$  in the statement.

*Proof.* Let  $\chi \in C_c^\infty(S)$  be a fixed probability measure supported in the ball  $\{s \in S : \|s\| \leq 2\}$ .

Denote by  $L_0^2(\nu)$  the orthogonal complement of the constant functions. By the spectral gap, furnished by Proposition 6.7.1, we have for  $F \in L_0^2(\nu_{x_0 S})$ :

$$\|F \star \chi\|_{L^2(\nu)} \leq (1 - \delta) \|F\|_{L^2(\nu)}$$

where  $\delta > 0$  depends only on  $\chi$  and the spectral gap for  $S$  acting on  $L^2(\nu_{x_0 S})$ . In view of the statement of Proposition 6.7.1, which guarantees such uniform spectral gaps, we may regard (after having fixed a choice of  $\chi$  for each subgroup  $S$ )  $\delta$  as depending only on  $G, H$ .

Let  $\chi^{(1)} = \chi$  and  $\chi^{(n)} = \chi^{(n-1)} \star \chi$  for  $n$  a positive integer. Here  $\star$  denotes convolution on  $S$ . By the definition of “almost invariance” (§3.10), we have

$|\mu(F \star \chi) - \mu(F)| \leq \epsilon \mathcal{S}_d(F)$ . In consequence, there exists  $E_\chi > 1$  so that:

$$(15.2) \quad |\mu(F \star \chi^{(n)}) - \mu(F)| \ll \epsilon E_\chi^{nd} \mathcal{S}_d(F)$$

Now, by (3.7), we see that the fibers of  $x \mapsto xs$ , considered as a map  $\{s \in S : \|s\| \leq 2\} \rightarrow X$ , have size  $\ll \text{ht}(x)^{\dim(S)\kappa_3}$ ; thus

$$|F \star \chi(x)| \ll \text{ht}(x)^{\kappa_3 \dim S} \int |F| d\text{vol} \ll \text{vol}(x_0 S) \text{ht}(x)^{\kappa_3 \dim(S)} \|F\|_{L^2(\nu)}.$$

Applying this to  $F \star \chi^{(n-1)}$ , we obtain:

$$(15.3) \quad |F \star \chi^{(n)}(x)| \ll \text{vol}(x_0 S) \text{ht}(x)^{\kappa_3 \dim(S)} (1 - \delta)^n \|F\|_{L^\infty}.$$

We combine (15.2), (15.3), and the fact  $|F \star \chi^{(n)}(x)| \leq \|F\|_{L^\infty} \ll \mathcal{S}_d(F)$ . It results

$$\mu(F) \ll \mathcal{S}_d(F) \left( \epsilon E_\chi^{nd} + \int_{x_0 S} \min(1, \text{vol}(x_0 S) (1 - \delta)^n \text{ht}(x)^{\kappa_3 \dim(S)}) d\mu \right)$$

To estimate the bracketed quantity on the right-hand side, we split the integral into  $x_0 S \cap \mathfrak{S}(R_1)$  and its complement, where  $R_1$  is a parameter that will be optimized. In view of Lemma 3.6.1, the bracketed quantity on the right-hand side is bounded by

$$\epsilon E_\chi^{nd} + \text{vol}(x_0 S) (1 - \delta)^n R_1^{\kappa_3 \dim S} + c_8 R_1^{-\kappa_4}$$

We now choose the free parameters  $R_1$  and  $n$  so that the three quantities  $\epsilon E_\chi^{nd}$ ,  $\text{vol}(x_0 S) (1 - \delta)^n R_1^{\kappa_3 \dim S}$ , and  $R_1^{-\kappa_4}$  are of comparable size. The stated result follows.  $\square$

## 16. PROOF OF THEOREM 1.3

We shall say that  $\mu$  is  $[S, \epsilon, d_S]$ -almost invariant if  $\mu$  is  $\epsilon$ -almost invariant under a connected intermediate subgroup  $S$  w.r.t. a Sobolev norm  $\mathcal{S}_{d_S}$ .

The precise analog to the non-effective proof, or of §2.7, would be to choose a “maximal”  $S$  under which  $\mu$  is almost invariant. However, we find it clearer in this effective setting to present instead the proof by iteration, i.e. building up, dimension by dimension, a larger and larger  $S$  under which  $\mu$  is almost invariant.

### 16.1. Completion of the proof of Theorem 1.3.

**Lemma.** *Suppose that  $\mu$  is  $[S, \epsilon, d_S]$ -almost invariant. There exists constants  $c_2(d_S)$ ,  $\kappa_{18}$ ,  $\kappa_{19}$ ,  $\kappa_{20}$ , and  $d'_S$  so that for any  $\epsilon$  sufficiently small (i.e.  $\epsilon \ll 1$ ) either:*

- $|\mu(f) - \mu_{x_0 S}(f)| \leq c_1(d_S) \epsilon^{\kappa_{19}(d_S)} \mathcal{S}_{d_S}(f)$ , for some closed orbit  $x_0 S$  of volume  $\leq \epsilon^{-\kappa_{18}(d_S)}$ .
- The measure  $\mu$  is  $[S_*, c_2(d_S) \epsilon^{\kappa_{20}(d_S)}, d'_S]$  almost invariant, where the connected subgroup  $S_* \supset H$  has larger dimension than  $S$ .

Apply Proposition 14.2 with  $\zeta = \kappa_{16}/d_S$ . If it fails to be applicable, we are in the first case of the Lemma, by Proposition 15.1. Indeed, we may take  $\kappa_{18}(d_S) = \kappa_{16}/d_S$  and  $\kappa_{19}(d_S) = \kappa_{17}/d_S$ .

Otherwise, Proposition 14.2 produces  $d'_S > d_S$  and  $\xi$ , both of which depend only on  $G, H, d_S$ , and two points  $x_1, x_2$  so that  $x_1, x_2$  are both  $[\|r\|^{-\kappa_8}, \|r_1\|^{-\kappa_9}]$ -generic for  $\mu$  w.r.t  $\mathcal{S}_{d'_S}$  moreover  $\|r\| \leq \epsilon^\xi$ .

Now we apply Proposition 10.2 with  $\mu_1 = \mu_2 = \mu$ . It shows that  $\mu$  is  $\ll_{d_S} \epsilon^{\min(1/2, \kappa_7 \xi/2)}$ -almost invariant under an element  $Z \in \mathfrak{r}$  which satisfies  $\|Z\| = 1$ . In view of Proposition 8.1,  $\mu$  is  $\ll_{d_S} \epsilon^{\kappa_{13} \min(1/2, \kappa_7 \xi/2)}$ -almost invariant under a subgroup  $S_* \supset H$  of strictly larger dimension.

The proof of the Lemma is complete.

**16.2. Proof by iteration.** We now prove Theorem 1.3 by iterating the Lemma. Now, by definition,  $\mu$  is  $[H, \epsilon, d_H]$ -almost invariant for arbitrary  $\epsilon$  and  $d_H = \kappa_5 + 1$ . By the lemma, for small enough  $\epsilon$  either:

- (1)  $|\mu(f) - \mu_{x_0 H}(f)| \ll \epsilon^{\kappa_{19}(d_H)} \mathcal{S}_{d_H}(f)$ , and  $x_0 H$  has volume  $\leq \epsilon^{-\kappa_{18}(d_H)}$ .<sup>29</sup>
- (2)  $\mu$  is  $[S_1, c_1 \epsilon^{\xi_1}, d_1]$ -almost invariant for some  $S_1 \supset H$  of strictly larger dimension; here  $\xi_1, d_1$  depend only on  $H, G$  and  $c_1 \ll 1$ .

Suppose the second case occurs. By the Lemma again, applied with  $c_1 \epsilon^{\xi_1}$  instead of  $\epsilon$ , one of the following occur for small enough  $\epsilon$ :

- (1)  $|\mu(f) - \mu_{x_0 S_1}(f)| \ll \epsilon^{\xi_1 \kappa_{19}(d_1)} \mathcal{S}_{d_1}(f)$ , where  $x_0 S_1$  is a closed  $S_1$ -orbit of volume  $\ll \epsilon^{-\xi_1 \kappa_{18}(d_1)}$ , or:
- (2)  $\mu$  is  $[S_2, c_2 \epsilon^{\xi_2}, d_2]$ -almost invariant for some  $S_2 \supset H$  of strictly larger dimension than  $S_1$ . Here  $\xi_2, d_2$  depends only on  $H$  and  $G$ , whereas  $c_2 \ll 1$ .

Iterating this process – which we can do at most  $\dim(G) - 1$  times – we arrive at the following conclusion:

$$|\mu(f) - \mu_{x_0 S_j}(f)| \ll \epsilon^{\xi_j \kappa_{19}(d_j)} \mathcal{S}_{d_j}(f)$$

where  $S_j \supset H$  has dimension  $\geq \dim(H) + j$ , and  $x_0 S_j$  is a closed orbit of volume  $\ll \epsilon^{-\xi_j \kappa_{18}(d_j)}$ .

In the rest of the proof, we will abbreviate an expression like  $|\mu(f) - \mu_{x_0 S}(f)| \leq \epsilon \mathcal{S}_d(f)$  to the phrase “ $\mu$  and  $\mu_{x_0 S}$  are  $\epsilon$ -close w.r.t.  $\mathcal{S}_d$ ,” sometimes we will suppress mention of the Sobolev norm.

Let  $\Delta$  be the supremum of all quantities  $\xi_j \kappa_{18}(d_j)$  that may arise through the above process. Let  $\delta$  be the infimum of all quantities  $\xi_j \kappa_{19}(d_j)$  that may arise through the above process. Let  $d$  be the supremum of all  $d_j$  that may arise through the above process. These three constants depend only on  $H$  and  $G$ , for there are a finite list of chains of intermediate subgroups  $H \subset S_1 \subset S_2 \subset \dots S_j$ .

Thus, we conclude that for  $\epsilon$  smaller than some constant  $\epsilon_0 \ll 1$ , there is an intermediate subgroup  $S_j \supset H$  and a closed  $S_j$ -orbit  $x_0 S_j$  of volume

<sup>29</sup>In fact, in the case,  $\mu = \mu_{x_0 H}$ ; we write it in a way that will resemble the other steps of the iteration.

$\leq c_3\epsilon^{-\Delta}$ , so that  $\mu$  is  $\leq c_4\epsilon^\delta$ -close w.r.t.  $\mathcal{S}_d$  to the  $S_j$ -invariant probability measure on  $x_0S_j$ .

The conclusion of our theorem follows: choose  $\epsilon$  so that  $c_3\epsilon^{-\Delta} = V$ . Then the above applies for  $V \gg 1$ , showing that  $\mu$  is  $\ll V^{-\delta/\Delta}$ -close w.r.t.  $\mathcal{S}_d$  to the  $S$ -invariant probability measure on the closed orbit  $x_0S$  of volume  $\leq V$ , where  $S$  is some intermediate subgroup.

By choosing appropriately  $V_0$  (from the statement of Theorem 1.3) in a way depending on  $G$ ,  $H$  and  $\Gamma$ , we can remove the implicit constant by lowering the exponent:  $\mu$  is  $\leq V^{-\delta/(2\Delta)}$ -close to the measure on  $x_0S$  whenever  $V \geq V_0$ .  $\square$

### 16.3. Proof of the topological Theorem 1.4.

*Proof of Theorem 1.4.* Notations as in Theorem 1.3; especially, let  $\delta$ ,  $d$ , and  $V_0$  be as in that Theorem. Let  $N$  be a large integer; it will be chosen in course of the proof to depend only on  $G, H$ .

Let  $\mathfrak{J}$  equal the number of intermediate subgroups  $H \subseteq S \subseteq G$ ; it is finite by Lemma 3.4.1.

Consider the set  $\mathcal{V}$  of all volumes of closed  $S$ -orbits  $x_0S$ , where  $H \subset S \subset G$ . The cardinality of  $\mathcal{V}$  is  $\leq \mathfrak{J}$ .

Take  $V > V_0^{N^{\mathfrak{J}+1}}$  and consider the intervals

$$(V^{1/N^{\mathfrak{J}+1}}, V^{1/N^{\mathfrak{J}}}], \dots, (V^{1/N^d}, V^{1/N^{d-1}}], \dots, (V^{1/N}, V], \text{ for } 1 \leq d \leq \mathfrak{J} + 1.$$

It is clear that one of these intervals will contain no element of  $\mathcal{V}$ . Call it  $(X, X^N]$ . We apply Theorem 1.3 with the parameter “ $V$ ” set to  $X^N$ . It follows that there exists a closed  $S$ -orbit  $x_0S$  of volume  $\leq X$ , for some intermediate subgroup  $H \subset S \subset G$ , so that:

$$(16.1) \quad \left| \int f d\mu - \int f d\mu_{x_0S} \right| < X^{-N\delta} \mathcal{S}_d(f),$$

Now, let  $x \in x_0S$ . Let  $\eta, f_x$  be as in the proof of Lemma 10.3.1; roughly  $\eta$  is any number less than the injectivity radius at  $x$ , and  $f_x$  is a bump function around  $x$  of radius  $\eta$ . The Sobolev norms of  $f_x$  are estimated in (10.7).

Applying (16.1), we see that:

$$|\mu(f_x) - \mu_{x_0S}(f_x)| \ll X^{-N\delta} \text{ht}(x)^d \eta^{-d}.$$

On the other hand, it is clear that

$$\mu_{x_0S}(f_x) \gg \eta^{\dim(S)} \text{vol}(x_0S)^{-1} \gg \eta^{\dim(S)} X^{-1}.$$

In particular, so long as:

$$\eta^{\dim(S)+d} \gg X^{1-N\delta} \text{ht}(x)^d$$

we will have  $\mu(f_x) > 0$ ; in particular,  $\text{supp}(\mu)$  must intersect the  $\eta$ -ball around  $x$ .

Specialize to the case when  $\Gamma \backslash G$  is compact; the general case of Theorem 1.4 follows in a precisely similar way. Taking  $N$  sufficiently large, applying

Lemma 10.3.1, and noting that  $X \geq V^{1/N^{2+1}}$ , we conclude that  $\text{supp}(\mu)$  intersects any ball on  $x_0S$  of radius  $\geq V^{-*}$ . Here the notion of “ball” is taken with respect to the induced Riemannian metric on  $x_0S$ .  $\square$

From the proof we can extract the following Corollary. In essence it is equivalent to the Theorem, but it is a formulation that seems at first a little stronger and is helpful in various contexts.

**16.3.1. Corollary.** *Notation as in Theorem 1.3, let  $\Delta > 0$  be real. There exists  $r \in (0, 1)$  and  $W_0 > 0$ , depending on  $G, H, \Gamma$ , and  $\Delta$ , so that:*

*For any closed  $H$ -orbit  $x_0H$  and any  $W > W_0$  there exists  $V \in [W^r, W]$ , an intermediate subgroup  $S$ , and a closed orbit  $x_0S$  of volume  $\leq V$ , so that  $\mu$  is  $V^{-\Delta}$ -close to  $\mu_{x_0S}$  w.r.t.  $\mathcal{S}_d$ .*

In other terms, we have “amplified” the exponent  $\delta$  of Theorem 1.3 to an arbitrarily large  $\Delta$ . Of course, this carries a hidden cost.  $\square$

## 17. AN ARITHMETIC APPLICATION: DISTRIBUTION OF INTEGRAL POINTS ON PREHOMOGENEOUS HYPERSURFACES

### 17.1. Introduction.

**17.1.1. Discriminant and height.** Notation as in our main theorem, closed  $H$ -orbits also have an *arithmetic* invariant, the “discriminant” which measures their arithmetic complexity (cf. [18]). In arithmetic applications, what one can easily measure is the discriminant, rather than the volume, of a closed  $H$ -orbit; thus we shall present a proposition relating the two. This result is Proposition 17.4.

We have already indicated some arithmetic applications of our results in §1.8.2; mainly to see how Proposition 17.4 arises naturally, we shall present another class of applications: to problems of Linnik type.

**17.1.2. Linnik problems.** By a *Linnik-problem*, we have in mind the following:  $f$  is a homogeneous polynomial on a  $\mathbb{Q}$ -vector space, and we wish to analyze the distribution of integral points on the level set  $\{f^{-1}(d)\}$ . For a discussion of problems of this type, see also [44]. In the case when  $f$  is *prehomogeneous*, this problem is amenable to analysis by our methods (cf. [21]). We shall present a quantitative theorem in this direction, for certain classes of such  $f$ , in Proposition 17.6.

This class of applications builds on the work of others. In particular, the applicability of Ratner’s theorem to these problems was observed by A. Eskin and H. Oh [22], related problems were studied by W. Gan and H. Oh. The idea of using invariant theory to handle focussing problems originates in work of A. Yuzvinkov [68, 67].

**17.2. Some comments on heights.** Let  $W$  be a  $\mathbb{Q}$ -vector space equipped with a Euclidean norm and an integral lattice  $W_{\mathbb{Z}} \subset W$ . The *height* of a subspace  $W' \subset W$  is, by definition, the norm  $\|e_1 \wedge \cdots \wedge e_r\|_{\wedge^r W}$ , where

$e_1, \dots, e_r$  is a basis for  $W' \cap W_{\mathbb{Z}}$ , and the norm on  $\wedge^r W$  is that derived from  $W$ . In explicit terms:

$$\text{ht}(W')^2 = \det((e_i, e_j)).$$

We shall use the following simple principles:

(17.1) A subspace of low height has a basis of low height.

In explicit terms, one may choose a basis for  $W' \cap W_{\mathbb{Z}}$  so that  $\|e_i\| \ll \text{ht}(W')$ . This assertion is simply lattice reduction theory, together with the observation that the lengths of elements of  $W_{\mathbb{Z}}$  are bounded below.

(17.2)

A system of linear equations of low height has a solution set of low height,

In more explicit terms, given a  $m \times n$  integral matrix  $A$ , all of whose entries are bounded above by a constant  $\|A\|$ , the kernel of  $A$  – considered as a subspace of  $\mathbb{Q}^m$ , where we endow  $\mathbb{R}^m$  with the Euclidean norm – has height bounded by  $\ll \|A\|^{\kappa(m,n)}$ .

**17.3. The discriminant of a closed  $H$ -orbit.** Notation as in our main theorem. Suppose that  $\Gamma gH$  is a closed  $H$ -orbit. We shall attach to it an arithmetic invariant, the *discriminant*. The corresponding procedure when  $H$  is a torus was introduced in [18].

First of all, we observe that, with  $\Lambda_g = \Gamma \cap gHg^{-1}$ ,

(17.3) the Zariski closure of  $\Lambda_g$  has Lie algebra  $\text{Ad}(g)\mathfrak{h}$ .

This follows from the Borel-Wang density theorem [37, Chapter II, Corollary 4.4], together with the fact that the algebraic group underlying  $H$  has no compact factors. Thus,  $\text{Ad}(g)\mathfrak{h}$  is a  $\mathbb{Q}$ -subspace of  $\mathfrak{g}$ .

Let  $r = \dim(H)$ , and define  $V = (\wedge^r \mathfrak{g})^{\otimes 2}$ ,  $V_{\mathbb{Z}} = (\wedge^r \mathfrak{g}_{\mathbb{Z}})^{\otimes 2}$ . Choosing any  $\mathbb{Q}$ -basis  $e_1, e_2, \dots, e_d$  for the Lie algebra of  $\text{Ad}(g)\mathfrak{h}$ , we set  $v_{gH} = \frac{(e_1 \wedge e_2 \wedge \dots \wedge e_r)^{\otimes 2}}{\det(B(e_i, e_j))^{\otimes 2}}$ , where  $B$  is the Killing form<sup>30</sup> on  $\mathfrak{g}$ . Then  $v_{gH}$  is independent of the choice of basis  $(e_i)$  for  $\text{Ad}(g)\mathfrak{h}$ . We define:<sup>31</sup>

$$\text{disc}(\Gamma gH) = \min\{m \in \mathbb{Z} : v_{gH} \in m^{-1}V_{\mathbb{Z}}\}.$$

**17.4. Proposition.** *There exists  $\kappa_{27} > 0$  so that, for any  $x \in \Gamma \backslash G$  with  $xH$  closed:*

- (1) *There exists a representative  $g \in G$  for  $x$  so that the height of  $\text{Ad}(g)\mathfrak{h}$  is  $\asymp \text{disc}(xH)^{1/2}$ .*
- (2)  $\text{vol}(xH) \gg \text{disc}(xH)^{\kappa_{27}}$ .
- (3)  $\text{vol}(xH) \ll \text{disc}(xH)^{\kappa}$ .

<sup>30</sup>The restriction of  $B$  to  $\mathfrak{h}$  is nondegenerate. This statement may be verified at the level of complexifications. Choose a real form of  $\mathfrak{h}_{\mathbb{C}}$  which is compact. The restriction of  $B$  to this subalgebra is negative definite, whence the assertion.

<sup>31</sup>By virtue of the fact that  $\mathfrak{g}_{\mathbb{Z}}$  is  $\Gamma$ -stable, this definition is independent of the choice of  $g$ .

We will not prove the third assertion, because we do not need it, but in fact the proof is substantially easier than the second assertion.

*Proof.* (Sketch). We have seen (Lemma 3.6.1) that there exists a fixed compact set  $K \subset G$  whose projection to  $\Gamma \backslash G$  necessarily intersects  $xH$ . Take  $g \in K$  to be any representative for  $x$ .

For the first observation, we observe that both the restriction of the Killing form, and the chosen Euclidean structure on  $\mathfrak{g}$ , induce definite quadratic forms on  $\wedge^r(\text{Ad}(g)\mathfrak{h})$ , for each  $g \in G/H$ . Clearly, the ratio of these forms vary continuously. In particular, if  $g$  lies within a fixed compact subset of  $G/H$ , the height of  $\text{Ad}(g)\mathfrak{h}$  is, up to constants, comparable to the square root of the discriminant, whence the first conclusion.

There are multiple methods of proof for the second assertion. We indicate a proof using dynamical ideas.

We first claim an effective version of (17.3). Namely, there exist constants  $c_1, \kappa_{28}$  so that:

$$\begin{aligned} & \text{the Zariski closure } \mathfrak{L}' \text{ of the group generated by} \\ & \{ \lambda \in \Lambda_g : \|\lambda\| \leq c_1 \text{vol}(xH)^{\kappa_{28}} \} \text{ has Lie algebra } \text{Ad}(g)\mathfrak{h}. \end{aligned}$$

This can be established using results about lattice point counting, such as those from §12.

This being established, let  $T \geq c_1 \text{vol}(xH)^{\kappa_{28}}$ . Let  $F := \{ \lambda \in \Lambda : \|\lambda\| \leq T \}$ .

Let us fix a representation  $\mathbf{G}$  on a  $\mathbb{Q}$ -vector space  $W$  with the following property: if  $W_{\mathbb{R}}^H$  denotes the  $H$ -fixed vectors in  $W_{\mathbb{R}}$ , then the pointwise stabilizer of  $W_{\mathbb{R}}^H$  in  $G$  is precisely the Zariski-closure of  $H$  within  $G = \mathbf{G}(\mathbb{R})$ . It is possible to do this, by Chevalley's theorem and the fact that the Zariski-closure of  $H$  is connected and does not admit algebraic characters. In particular,

$$(17.4) \quad \{v \in \mathfrak{g} : v.W_{\mathbb{R}}^H = 0\} = \mathfrak{h}.$$

Fix, once and for all, a Euclidean norm on  $W_{\mathbb{R}}$  and an integral lattice  $W_{\mathbb{Z}}$ .

Consider the fixed subspace  $W^F$  for  $F$  inside  $W$ . We claim that  $W^F$  has “low height”, i.e. bounded by a power of  $\|g\|$  and  $\text{vol}(xH)$ , when considered as a subspace of  $W$ . Indeed, one may replace  $F$  by a subset  $F' \subset F$ , with cardinality bounded in terms of  $G$  alone, so that  $W^{F'} = W^F$ , by similar arguments to §13.6. The claim follows from (17.2).

Our definitions are so that the subgroup generated by  $F$  is Zariski-dense in  $gHg^{-1}$ ; it follows that  $W_{\mathbb{R}}^F$  equals  $g.W_{\mathbb{R}}^H$ .

Therefore, the subalgebra of  $\mathfrak{g}$  defined by:

$$(17.5) \quad \{v \in \mathfrak{g} : v.W^F = 0\}$$

also has “low height”, in the same sense. This follows from (17.1) and (17.2). By our assumption, the subalgebra defined by (17.5) coincides with the Lie algebra of  $\text{Ad}(g)\mathfrak{h}$ .

We have established that  $\text{Ad}(g)\mathfrak{h} \subset \mathfrak{g}$  has “low height” in terms of  $\text{vol}(xH)$ ; in view of the proof of the first assertion of the present Proposition, we are done.  $\square$

We observe that this Proposition gives immediately an alternate proof of Lemma 10.3.1.

**17.5. Application to Linnik problems.** We shall focus on the following setting: Let  $\mathbf{G}$  be a semisimple algebraic  $\mathbb{Q}$ -group which acts on a  $\mathbb{Q}$ -vector space  $V$ , which preserves a polynomial invariant  $f : V \rightarrow \mathbb{Q}$ , and so that (algebraically) each level set of  $f$  is a single  $\mathbf{G}$ -orbit. Let  $V_{\mathbb{Z}} \subset V$  be a lattice, and  $\Gamma$  a congruence subgroup of  $G := \mathbf{G}(\mathbb{R})$  that preserves  $V_{\mathbb{Z}}$ .

This is, roughly speaking, the setting for the (arithmetic) study of “prehomogeneous vector spaces.” We shall impose an additional condition: that the stabilizer of a generic point is semisimple and has finite centralizer in  $\mathbf{G}$ .

Here are two specific instances with these properties (more can be found by examining the tables of prehomogeneous vector spaces).

Case A. Fix an integer  $r \geq 3$ . Let  $\mathbf{G} = \text{SL}(r)$ ,  $\Gamma = \text{SL}(r, \mathbb{Z})$ , take  $V_{\mathbb{Z}}$  to consist of  $r \times r$  integral symmetric matrices, and  $f = \det$ .

Here  $\dim(V) = r(r+1)/2$ ,  $\deg(f) = r$ , and the stabilizer of a generic point is a form of the orthogonal group  $\text{SO}(r)$ .

Case B. Fix  $r \in \{7, 8\}$ .<sup>32</sup> Let  $\mathbf{G} = \text{SL}(r)$ ,  $\Gamma = \text{SL}(r, \mathbb{Z})$ , let  $V_{\mathbb{Z}}$  be the space of alternating trilinear forms on  $\mathbb{Z}^r$ , with integral values; and  $f$  is the *discriminant* (cf. §17.7.4).

Here, if  $r = 7$ , then  $\dim(V) = 35$ ,  $\deg(f) = 7$  and the stabilizer of a generic point is a form of  $G_2$ . If  $r = 8$ , then  $\dim(V) = 36$  and  $\deg(f) = 16$ , and the stabilizer of a generic point is a form of  $\text{SL}(3)$ .

In Case B, some questions in Diophantine geometry on  $V$ , analogous to the Oppenheim conjecture, were studied by ergodic methods by Yuzvinsky and collaborators ([68], [67]) and, indeed, we use a technique to handle “focussing” analogous to that of [68, 67].

The following Proposition gives a quantitative solution to Linnik’s distribution problem in the Cases A and B above. A non-effective version in Case A was established by Eskin and Oh, [21]. We observe that, in both these cases, the degree is “too large” relative to the dimension for the Hardy-Littlewood method to be applicable. (For instance, Case A includes the case of a (special) cubic form in five variables.)

**17.6. Proposition.** *Let  $(V, f)$  be as in Case A or Case B, defined above.*

*Let  $V_{\mathbb{R}}^{nc}$  be the open subset of  $V_{\mathbb{R}}$  comprising points whose stabilizers in  $\mathbf{G}(\mathbb{R})$  are noncompact.*

*Suppose  $\Omega \subset \{x \in V_{\mathbb{R}}^{nc} : f(x) = 1\}$  is compact with smooth boundary.*

---

<sup>32</sup>The space with  $r = 6$  would still be prehomogeneous; however, the stabilizer of a generic point has an infinite centralizer.

Let  $d \rightarrow \infty$  vary through integers with bounded square part<sup>33</sup> and so that  $f^{-1}(d) \cap V_{\mathbb{Z}} \neq \emptyset$ .

Then:

$$|\{M \in V_{\mathbb{Z}} : f(M) = d, \frac{M}{d^{1/\deg(f)}} \in \Omega\}| = C_d(\text{vol}(\Omega) + O_{\Omega}(d^{-\delta})).$$

Here  $\liminf_d \frac{\log C_d}{\log d} > 0$ .

The proof will be only sketched. Presumably  $\frac{\log C_d}{\log d} \rightarrow \frac{\dim(V) - \deg(f)}{\deg(f)}$ , but we do not establish this.<sup>34</sup> We intend to elaborate on this and other applications in the  $S$ -arithmetic sequel to this paper.

The restriction that  $\Omega \subset V^{nc}$  can presumably be removed by  $p$ -adic methods, e.g. [20]. For “Case A” it is clear that  $f$  takes on all integer values, and so we can dispense with the restriction  $f^{-1}(d) \cap V_{\mathbb{Z}}$  be nonempty; it would be nice to establish the exact range of values taken by  $f$  in “Case B.”

**17.7. Prehomogeneous vector spaces: generalities.** Let us proceed in the general setting enunciated at the start of §17.5.

The set of points in  $V - f^{-1}(0)$  whose stabilizer is a prescribed subgroup of  $\mathbf{G}$  is a finite set of lines, by virtue of the assumption that the generic stabilizer is semisimple and has finite centralizer.

The level set  $f^{-1}(1)$  is a union of a finite collection of  $G = \mathbf{G}(\mathbb{R})$ -orbits. Fix representatives  $x_i$  for each orbit, and let  $H_i$  be the stabilizer of  $x_i$ . (We do not assume that  $H_i$  is connected; we will not go into details about the easy arguments required, in what follows, to get around connectedness issues.) We fix also a compact subset  $\Omega \subset f^{-1}(1)$ .

For  $d > 0$ , each level set  $f^{-1}(d)$  is – by scaling – identified with  $\bigcup_i G/H_i$ . For  $y \in f^{-1}(d) \cap V_{\mathbb{Z}}$ , let  $\bar{y}$  be its projection to  $f^{-1}(1)$ ; we may choose  $i(y), g_y$  so that  $\bar{y} = g_y \cdot x_{i(y)}$ .

The orbit  $\Gamma g_y H_i$  is then closed, because the group  $g_y H_i g_y^{-1}$  is the stabilizer of  $y$  and therefore corresponds to the real points of a  $\mathbb{Q}$ -algebraic group.

**17.7.1. Lemma.** *There exists  $c > 0$  with the following property: for any  $y$  with  $\bar{y} \in \Omega$  and  $f(y) \neq 0$ , the discriminant of  $\Gamma g_y H_i$  is  $\gg_{\Omega} \text{ht}(\mathbb{Q}.y)^c$ .*

*Proof.* Consider the fixed points for the Lie algebra  $\text{Ad}(g_y)\text{Lie}(H_i)$ .

This set of fixed points is a linear subspace; on the other hand, it intersects the Zariski-open set  $V - f^{-1}(0)$  in a nonempty finite set of lines. Therefore, it must consist of a single line. We may therefore characterize  $\mathbb{Q}.y$  as the fixed line for  $\text{Ad}(g_y)\text{Lie}(H_i)$ . (17.1) and (17.2) imply that the height of  $\mathbb{Q}.y$  is bounded by a power of the height of  $\text{Ad}(g_y)\text{Lie}(H_i)$ .

<sup>33</sup>I.e. the set of perfect squares which divide some  $d$  is bounded

<sup>34</sup>It is related to the question of the precise relationship between discriminant and volume of a periodic orbit. In principle, this can be reduced to the computation of Tamagawa number, but the local computations seem difficult in the most general setting.

Because  $\bar{y} \in \Omega$ , we may suppose that  $g_y$  belongs to a fixed compact subset within  $G$ . Our claim now follows from the first assertion of Proposition 17.4.  $\square$

17.7.2. *Proof of Proposition 17.6 in Case A.* Let us restrict to Case A.

Suppose that  $y \in V_{\mathbb{Z}}^{nc} \cap f^{-1}(d)$ , with  $\bar{y} \in \Omega$ .

It follows from Lemma 17.7.1 and the assumption on  $d$  (bounded square part) that:

$$(17.6) \quad \text{vol}(\Gamma g_y H_i) \gg d^{\kappa},$$

for some  $\kappa > 0$ .

Indeed, if we let  $\spadesuit$  be the g.c.d. of the entries of  $y$ , i.e. the largest integer so that  $\spadesuit^{-1}y \in V_{\mathbb{Z}}$ , then  $\spadesuit^r$  divides  $\det(y) = d$ . Because of the assumption that  $d$  had bounded square part,  $\spadesuit$  is also bounded. But the height of  $\mathbb{Q}.y$  is, up to bounded multiples, the norm of  $\spadesuit^{-1}y$ , whence our conclusion.

In combination with our main theorem, this implies Proposition 17.6 in “Case A.” See [21] for details of translating equidistribution of  $H$ -orbits to statements in the style of Proposition 17.6.

The idea, in words, is that our main theorem shows, in an effective sense, the equidistribution of  $\Gamma g_y H_i \subset \Gamma \backslash G$ . (We know, by (17.6), that this orbit has large volume). This, however, is equivalent to the uniform distribution (interpreted suitably) of  $\Gamma g_y$  on  $G/H$ , or, equivalent to the uniform distribution of the  $\Gamma$ -orbit  $\Gamma y$  on  $f^{-1}(d)$ . The integral points on  $f^{-1}(d) \cap V_{\mathbb{Z}}$  are the union of finitely many such orbits, whence our conclusion.

17.7.3. *Preparations for the proof in Case B.* In order to establish Proposition 17.6 in “Case B,” we shall give a little more background on alternating trilinear forms.

Let us recall that  $V_{\mathbb{Z}}$  is the space of *trilinear alternating tensors* on  $\mathbb{Z}^r$ . In explicit terms, an element  $t \in V$  is an alternating, trilinear map:

$$t : \mathbb{Z}^r \times \mathbb{Z}^r \times \mathbb{Z}^r \rightarrow \mathbb{Z}.$$

17.7.4. *Invariants and covariants.* For  $6 \leq r \leq 8$ , trilinear tensors have a *discriminant*: a polynomial function  $\text{disc} : V \rightarrow \mathbb{Q}$ , homogeneous of degree 4 (resp. 7, 16) when  $r = 6$  (resp. 7, 8). For definitions, we refer to [68, 67]; unfortunately, we do not know of any entirely simple definition.

There exists a polynomial  $\text{SL}(d)$ -equivariant map:

$$\text{cov} : V \rightarrow \text{Sym}^2 \mathbb{Q}^r$$

of degree 3 (when  $r = 7$ ) and of degree 10 (when  $r = 8$ ). This is proven in [68, 1.16] and [67, Definition 2.13].

The determinant of  $\text{cov}(t)$  is a scalar multiple of  $\text{disc}(t)^3$  ( $r = 7$ ) and a scalar multiple of  $\text{disc}(t)^5$  (when  $r = 8$ ).

By “clearing denominators”, we may suppose that  $\text{disc}$  maps  $V_{\mathbb{Z}}$  into  $\mathbb{Z}$ , and that  $\text{cov}$  maps  $V_{\mathbb{Z}}$  into  $\text{Sym}^2 \mathbb{Z}^r$ .

17.7.5. *Proof of Proposition 17.6 in Case B.* These preliminaries being established, let us turn to the proof of Proposition 17.6 in the case at hand.

We follow the general notations of §17.7, with  $\mathbf{G} = \mathrm{SL}_r$ ,  $V$  as above,  $f = \mathrm{disc}$ . In particular, we choose points  $x_i \in f^{-1}(1)$  as in §17.7. Let  $S_i$  be the stabilizer of  $\mathrm{cov}(x_i)$  inside  $G$ ; it is a special orthogonal group, which contains  $H_i$  by the equivariance of  $\mathrm{cov}$ .

Take  $y \in V_{\mathbb{Z}} \cap f^{-1}(d)$  so that the stabilizer of  $y$  is non-compact. Then  $\Gamma g_y S_{i(y)}$  is closed; for, by equivariance,  $\mathrm{Ad}(g_y) S_{i(y)}$  is the stabilizer of the integer vector  $\mathrm{cov}(y_i) \in \mathrm{Sym}^2 \mathbb{Z}^r$ , and therefore the real points of a  $\mathbb{Q}$ -subgroup.

It follows from the Lemma of §17.7 that the volumes of  $\Gamma g_y H_{i(y)}$  and  $\Gamma g_y S_{i(y)}$  are bounded below in terms of the respective heights of the lines  $\mathbb{Q}.y$  and  $\mathbb{Q}.\mathrm{cov}(y)$ . Now  $\mathrm{disc}(y) = d$  and  $\det(\mathrm{cov}(y))$  is a fixed multiple of  $d^3$  or  $d^{10}$ , according to whether  $r = 7$  or  $8$ . Reasoning as in §17.7.2, and we conclude that the height of  $\mathbb{Q}.y$  exceeds  $d^{1/\deg(f)}$ , and that the height of  $\mathbb{Q}.\mathrm{cov}(y)$  exceeds  $d^{3/7}$  resp.  $d^{1/8}$ , according to whether  $r = 7$  or  $r = 8$ .

Finally, it is proven in [68, 67] that the Lie algebra of  $S_i$  is the only such algebra intermediate between  $H_i$  and  $G$ .

The claimed result of Prop. 17.6 follows from Theorem 1.3. Indeed, our discussion above has shown that both  $\Gamma g_y H_{i(y)}$  and  $\Gamma g_y S_{i(y)}$  has “large” volume; applying the Theorem shows that  $\Gamma g_y H_{i(y)}$  is approximately uniformly distributed in  $\Gamma \backslash G$ . This translates into the statement of Proposition 17.6, as carried out in [21] and recalled in approximate form in §17.7.2.  $\square$

## APPENDIX A. PROOF OF LEMMA 3.4.1

In this section,  $k$  will be a local field of characteristic zero,  $\mathbf{G}$  a semisimple algebraic group over  $k$ ,  $\mathfrak{g}$  its Lie algebra,  $\mathfrak{h} \subset \mathfrak{g}$  a semisimple subalgebra.

We establish some preliminary results.

### A.1. Embeddings of semisimple Lie algebras.

**A.1.1. Lemma.** *Let  $\mathfrak{s}$  be a semisimple Lie algebra over  $k$ . There exist finitely many embeddings of  $\mathfrak{s}$  into  $\mathfrak{g}$ , up to  $\mathbf{G}(k)$ -conjugacy.*

*Proof.* It suffices to prove this statement over the algebraic closure  $\bar{k}$ . Indeed, that being assumed, the affine variety  $\mathbf{X}$  parameterizing embeddings of  $\mathfrak{s}$  into  $\mathfrak{g}$  is a finite union of homogeneous  $\mathbf{G}$ -spaces. The finiteness of Galois cohomology over local fields assures that, given a homogeneous  $\mathbf{G}$ -space  $\mathbf{Y}$ , the set of  $\mathbf{G}(k)$ -orbits on  $\mathbf{Y}(k)$  is also finite.

For the statement over  $\bar{k}$ , choose an embedding of  $\mathbf{G}$  into the general linear group  $\mathrm{GL}(n)$ . The representation theory of semisimple Lie algebras assures that the number of  $\mathrm{GL}(n, \bar{k})$ -orbits on homomorphisms  $\mathfrak{s} \rightarrow \mathfrak{gl}_n$  is finite. Our assertion then follows from [59, Theorem 7.1].  $\square$

### A.2. Parabolic subgroups.

**A.2.1. Lemma.** *If  $\mathbf{S}$  is a proper algebraic subgroup of  $\mathbf{G}$ , so that  $\mathbf{S} = N_G(R_u(\mathbf{S}))$ , then  $\mathbf{S}$  is parabolic.*

See [3] or [63].

**A.2.2. Lemma.** *Any algebraic subgroup of  $\mathbf{G}$ , with a nontrivial unipotent radical, is contained in a parabolic subgroup of  $\mathbf{G}$ .*

*Proof.* Let  $\mathbf{S}$  be an algebraic subgroup with nontrivial unipotent radical. Define, inductively,  $\mathbf{S}^{(0)} = \mathbf{S}$  and  $\mathbf{S}^{(j+1)} = N_G(R_u(\mathbf{S}^{(j)}))$ , for  $j \geq 0$ . Note that  $\mathbf{S}^{(j)}$  normalizes  $R_u(\mathbf{S}^{(j)})$  so that  $\mathbf{S}^{(j)} \subset \mathbf{S}^{(j+1)}$ . Furthermore,  $R_u(\mathbf{S}^{(j)}) \subset R_u(\mathbf{S}^{(j+1)})$ . Since unipotent groups are connected in characteristic zero, the increasing chain  $R_u(\mathbf{S}^{(j)})$  of algebraic subgroups must necessarily stabilize. Therefore, the same must hold for the chain  $\mathbf{S}^{(j)}$ . The previous lemma allows us to conclude the proof.  $\square$

### A.3. Intermediate subgroups.

**A.3.1. Lemma.** *Suppose  $\mathfrak{h} \subset \mathfrak{g}$  is semisimple and has trivial centralizer. Then any intermediate subalgebra is semisimple.*

*Proof.* Take  $\mathfrak{h} \subset \mathfrak{s} \subset \mathfrak{g}$ . Suppose  $\mathfrak{s}$  were not semisimple. Then  $\mathbf{S}$ , the connected component of the normalizer of  $\mathfrak{s}$ , must also fail to be semisimple.

It suffices, therefore, to prove that any intermediate algebraic subgroup between  $\mathbf{H}$  and  $\mathbf{G}$  is semisimple. Here  $\mathbf{H}$  is the connected algebraic group with Lie algebra  $\mathfrak{h}$ .

We claim that  $\mathbf{S}$  cannot have nontrivial unipotent radical. In view of Lemma A.2.2, were this false, then  $\mathbf{H}$  would be contained in a proper parabolic subgroup  $\mathbf{P}$  of  $\mathbf{G}$ . Consider the representation of  $\mathbf{H}$  on the Lie algebra  $\mathfrak{p}$ ; let  $\mathfrak{n}$  be the Lie algebra of the unipotent radical  $\mathbf{N}$  of  $\mathbf{P}$ . The quotient group  $\mathbf{P}/\mathbf{N}$  has nontrivial center; therefore,  $\mathbf{H}$  fixes a subspace in its adjoint action on  $\mathfrak{p}/\mathfrak{n}$ . By semisimplicity of  $\mathbf{H}$ , it also fixes a subspace in its action on  $\mathfrak{p}$ . Therefore,  $\mathfrak{h}$  has a nontrivial centralizer, contradiction.

To conclude note that in absence of a unipotent radical, the radical of the subgroup  $\mathbf{S}$  is central. Since  $\mathfrak{h}$  has trivial centralizer,  $\mathbf{H}$  has finite centralizer. Therefore, we see that the radical of  $\mathbf{S}$  must be trivial.  $\square$

**A.3.2. Lemma.** *Suppose  $\mathfrak{h} \subset \mathfrak{g}$  is semisimple and has trivial centralizer. Then there exist only finitely many intermediate subalgebras  $\mathfrak{h} \subset \mathfrak{s} \subset \mathfrak{g}$ .*

*Proof.* It suffices to prove this statement over the algebraic closure.

There exist only finitely many isomorphism classes of semisimple Lie algebras  $\mathfrak{s}$  that can be embedded into  $\mathfrak{g}$ . For each such, there exist – by Lemma A.1.1 – only finitely many  $\mathbf{G}(\bar{k})$ -conjugacy classes of embeddings. Choose representatives for the image of every such embedding, calling them  $\mathfrak{s}_1, \dots, \mathfrak{s}_m$ . Let  $\mathbf{S}_j \subset \mathbf{G}$  be a connected semisimple algebraic group with Lie algebra  $\mathfrak{s}_j$ .

Then, for each  $j$ , there exist but finitely many  $\mathbf{S}_j(\bar{k})$ -conjugacy classes of subalgebras of  $\mathfrak{s}_j$ , isomorphic to  $\mathfrak{h}$ . Call them  $\mathfrak{h}_{ij}$ ,  $1 \leq i \leq N_j$ .

Take any intermediate subalgebra  $\mathfrak{s} \supset \mathfrak{h}$ . The pair  $(\mathfrak{h} \subset \mathfrak{s})$  is conjugate under  $\mathbf{G}(\bar{k})$  to  $(\mathfrak{h}_{ij} \subset \mathfrak{s}_j)$  for some  $j$  and  $i$ .

Thus, there exist only finitely many possibilities for the conjugacy class of the pair  $(\mathfrak{h} \subset \mathfrak{s})$ . However, the normalizer of  $\mathfrak{h}$  in  $\mathbf{G}$ , contains  $\mathbf{H}$  as a finite index subgroup. The claimed result follows.  $\square$

*Proof of Lemma 3.4.1.* The results of this section imply all the claims, save that every intermediate subgroup  $S \supset H$  has no compact factors. However,  $\mathfrak{h}$  would, by necessity, centralize the Lie algebra of any compact factor. This contradicts the assumption that  $\mathfrak{h}$  has no center.  $\square$

## APPENDIX B. PROOF OF LEMMA 3.6.1.

For the proof of Lemma 3.6.1 we will use the non-divergence results for actions of unipotent subgroups.

In the paper [36], by G.M., it has been shown that a point of low height “returns to the set of low height infinitely often” under a one-parameter unipotent flow.

Dani has refined this in [10] to show that the corresponding conclusion remains valid, even without the constraint that the original point have low height, unless there is a rational subspace that is invariant under the unipotent one-parameter subgroup considered (i.e. a rational constraint prohibits that).

This is almost what we need to prove (1) of Lemma 3.6.1. What we will use and actually need for (2) of Lemma 3.6.1 is a quantification of this phenomenon. We use results from the paper [27] by Kleinbock and G.M.

*Proof.* For  $g \in G$ , we call a subspace  $V \subset \mathfrak{g}$  *g-rational* if  $V \cap \text{Ad}_g^{-1}\mathfrak{g}_{\mathbb{Z}}$  is a lattice in  $V$ . We define the covolume of a  $g$ -rational subspace to be the volume of  $V/(V \cap (\text{Ad}_g^{-1}\mathfrak{g}_{\mathbb{Z}}))$ . A  $H$ -invariant,  $g$ -rational subspace  $V$  of low covolume prohibits a point of low height on  $\Gamma gH$ . (Let us note that  $\text{Ad}_h : V \rightarrow V$  has determinant 1, because  $H$  is semisimple).

We shall show that, no matter what  $g$  is, there are *no*  $g$ -rational subspaces of low covolume. This will be by induction on  $\dim(V)$ . The notion of “low covolume” will be specified as we go along. After this is done, we may establish statement (1) of the Lemma.

The beginning of the induction is rather trivial; there are no  $H$ -invariant lines  $V \subset \mathfrak{g}$  by our assumption that the centralizer of  $H$  on  $\mathfrak{g}$  is trivial.

Suppose now we have already established that there are no  $H$ -invariant  $g$ -rational subspaces of covolume less than  $c < 1$  and dimension  $< k$  for some  $k \leq \dim G$ . Suppose also that  $V$  is an  $H$ -invariant  $g$ -rational subspace with covolume  $v$  and dimension  $k$  for some  $v > 0$ . As a first step towards our inductive step:

*Claim.* If  $v$  is sufficiently small, there exists  $h \in H$  so  $\text{Ad}_h \text{Ad}_g^{-1}\mathfrak{g}_{\mathbb{Z}} \cap V$  has a basis consisting of vectors of length  $\ll v^{\frac{1}{k}}$ .

To prove the claim we shall use the result of [27] mentioned. There are only finitely many sublattices  $L_i \subset \text{Ad}_g^{-1} \mathfrak{g}_{\mathbb{Z}} \cap V$  of covolume less than  $c$  and dimension strictly less than  $k$ . Let  $U$  be the image of  $u(t)$ , defined in (3.1); consider the set of  $h \in H$  s.t.  $\text{Ad}(h)U$  preserves each  $L_i$ . This is an algebraic condition, i.e. the real points of a real algebraic subvariety of  $\mathbf{G}_{\mathbb{R}}$ . Thus either:

- (1) There is  $h \in H$  such that  $hUh^{-1}$  does not leave any of these subspaces invariant;
- (2) There exists a sublattice  $L_j$  so that we have that  $hUh^{-1}(L_j) = L_j$  for all  $h \in H$ .

The normal subgroup generated by  $U$  must coincide with  $H$ . Thus, in the second case,  $H$  preserves  $L_j$ , a contradiction. We are thereby in the first case; conjugating  $U$ , we may assume without loss of generality that  $U$  leaves invariant no  $g$ -rational subspace  $W \subset V$  of dimension  $< k$  and covolume  $< c$ .

Put  $h(t) = \text{Ad}_{u(t)}|_V g' \in \text{SL}(V)$ , a polynomial in  $t$ . The same holds for  $\bigwedge^\ell h(t) \in \text{SL}(\bigwedge^\ell V)$  for any  $\ell < k$ . Therefore, the square of  $\psi_W(t) = \text{covol}(h(t)(W \cap \text{Ad}_g^{-1} \mathfrak{g}_{\mathbb{Z}}))$  is also a polynomial for any subspace  $W \subset V$  of dimension  $\ell < k$  which intersects  $\text{Ad}_g^{-1} \mathfrak{g}_{\mathbb{Z}}$  in a lattice.<sup>35</sup>

There are only finitely many subspaces  $W \subset V$  of dimension  $\ell < k$  with  $\psi_W(0) < cv^{-\ell/k}$ . For such  $W$ , it is possible to choose some  $t_W$  with  $\psi_W(t_W) \geq cv^{-\ell/k}$ . I.e. for some  $r > 0$  we know that if  $v$  is sufficiently small depending on  $c$  and  $k$  that the supremum norm of these functions satisfy  $\|\psi_W\|_{[-r,r]} \geq 1$  for all such subspaces  $W$ .

By [27, Prop. 3.2] the polynomials  $\psi_W^2$  are all uniformly  $(C, \alpha)$ -good for some  $C > 0$  and  $\alpha > 0$  that depend only the degree of  $h(t)$  and so on  $\dim G$  — the same holds for  $\psi_W$  (with slightly different constants). We do not need to define this notion since this is only used as the second and last assumption of [27, Thm. 5.2]. From this quantitative non-divergence theorem we conclude that

$$|\{t \in [-r, r] : h(t)(\text{Ad}_{g^{-1}} \mathfrak{g}_{\mathbb{Z}} \cap V) \text{ contains an element of size } < \epsilon\}| \ll \epsilon^\alpha r,$$

where the implicit constant depends on  $C$  and  $k$ . If we choose  $\epsilon$  small enough in comparison to the implicit constant we see that there exists some  $t$  such that  $h(t)(\text{Ad}_{g^{-1}} \mathfrak{g}_{\mathbb{Z}} \cap V)$  does not contain an element of size smaller than  $\epsilon$ . Going back to  $\mathfrak{g}$  this shows that  $V \cap (\text{Ad}_{u(t)g^{-1}} \mathfrak{g}_{\mathbb{Z}})$  does not contain an element of size  $< \epsilon v^{1/k} \ll v^{1/k}$ . Therefore,  $V \cap (\text{Ad}_{u(t)g^{-1}} \mathfrak{g}_{\mathbb{Z}})$  is generated by elements of size  $\ll v^{1/k}$  by Minkowski's theorem on successive minima for lattices and since  $v$  is the covolume.

To conclude the induction we need to show that there cannot be any  $H$ -invariant  $g$ -rational subspace  $V$  which has small covolume  $v$ . By the above

<sup>35</sup>Also note that  $\psi_W$  agrees up to a bounded multiplicative factor with the function  $\psi_{W \cap \mathfrak{g}^{-1} \mathfrak{g}_{\mathbb{Z}}}$  defined in [27, §5] which is defined by a different norm on  $\bigwedge^{\dim W} \mathbb{R}^k$ .

claim we may assume that  $V \cap (\text{Ad}_g^{-1} \mathfrak{g}_{\mathbb{Z}})$  is generated by elements  $\{w_i\}$  of size  $\ll v^{1/k}$ . Recall that for any  $w, w' \in \mathfrak{g}$  we have  $\|[w, w']\| \leq \|w\| \|w'\|$ . Suppose  $v$  is small enough such that  $V \cap (\text{Ad}_g^{-1} \mathfrak{g}_{\mathbb{Z}})$  is generated by elements of size  $< \frac{1}{2}$ . (This choice of  $v$  depends only on the implicit constants in the argument above).

Therefore,  $[w_i, w_j]$  have length  $\leq \frac{1}{4}$ ; similarly for higher order commutators. Also recall that  $\mathfrak{g}_{\mathbb{Z}}$  satisfies  $[\mathfrak{g}_{\mathbb{Z}}, \mathfrak{g}_{\mathbb{Z}}] \subset \mathfrak{g}_{\mathbb{Z}}$  and so all of these commutators belong to the lattice  $\text{Ad}_g^{-1} \mathfrak{g}_{\mathbb{Z}}$ . Therefore, the Lie algebra  $\mathfrak{l}$  generated by  $V$  is nilpotent.

Moreover, since  $V$  is assumed to be invariant under  $H$ , the same holds for the Lie algebra  $\mathfrak{l}$ . Since  $H$  is semisimple, we must have that the Lie algebra  $\mathfrak{h}$  of  $H$  and  $\mathfrak{l}$  intersect trivially since otherwise we have found a nilpotent Lie ideal  $\mathfrak{l} \cap \mathfrak{h}$  contained in  $\mathfrak{l}$ . This shows that there is an intermediate subgroup  $S \subset G$  with Lie algebra  $\mathfrak{h} \otimes \mathfrak{l}$  which fails to be semisimple in contradiction to Lemma 3.4.1.

The proof of (1) is now similar to part of the above induction. In fact we have established that there are no  $H$ -invariant  $g$ -rational subspaces  $V$  of small covolume for any dimension  $k < \dim G$ , and as above we may assume this is also true for  $U$ -invariant  $g$ -rational subspaces. Therefore, the square of  $\psi_V(t) = \text{covol}(\text{Ad}_{u(t)}(V \cap \text{Ad}_g^{-1} \mathfrak{g}_{\mathbb{Z}}))$  (for any  $g$ -rational subspace  $V$ ) is either an unbounded polynomial or equal to a constant  $\geq \rho \gg 1$ . Therefore, for any large enough  $r > 0$  we will have  $\|\psi_W\|_{[-r, r]} \geq \rho$  for all  $g$ -rational subspaces, and so

$$(B.1) \quad |\{t \in [-r, r] : \Gamma g u(t) \notin \mathfrak{S}(\epsilon^{-1})\}| \ll \left(\frac{\epsilon}{\rho}\right)^{\alpha} r,$$

again by [27, Thm. 5.2]. Here the implicit constant only depends on  $G$ . If we choose  $\epsilon = R_0^{-1}$  small enough, then for some  $t$  we have  $\Gamma g u(t) \in \mathfrak{S}(R_0)$  as claimed.

Turning to (2) assume  $\mu$  is  $H$ -invariant and  $H$ -ergodic, then by the Mautner phenomenon  $U$  acts also ergodically. Therefore, by the pointwise ergodic theorem we can find some  $x \in X$  for which the ergodic averages along  $u(t)$  for the characteristic functions of the sets  $\mathfrak{S}(n)$  converge to  $\mu(\mathfrak{S}(n))$  for all integers  $n \geq 1$ . Combined with (B.1) this gives

$$\mu(X \setminus \mathfrak{S}(n)) = \lim_{r \rightarrow \infty} \frac{1}{2r} \int_{-r}^r \chi_{X \setminus \mathfrak{S}(n)}(xu(t)) dt \ll n^{-\alpha}$$

as required.  $\square$

## APPENDIX C. PROOF OF (6.1): UNIFORM SPECTRAL GAP IMPLIES $L^p$ COEFFICIENTS.

This section is devoted to a discussion of how to extract (6.1) from statements in the literature; we thank Erez Lapid for his help with the details. We apologize to the reader for the piecemeal nature of this proof.

Essentially we make use of a simple “quantization” phenomenon: any point in the unitary dual of a semisimple Lie group, at which the behavior of irreducible representations “changes,” is restricted to sets where certain parameters are restricted to be integral. In particular, away from the identity representation, such “changes in behavior” can only occur at certain discrete points; in combination with the Howe-Moore theorem, this will establish the result.

See [8, Theorem 2.4.2] for groups of real rank  $\geq 2$  and [8, Theorem 2.5.2] for the real rank one groups with property  $T$ . We shall therefore discuss carefully the rank one groups that fail to have property  $T$ , namely,  $\mathrm{SO}(n, 1)$  or  $\mathrm{SU}(n, 1)$ .

We use Theorem 6 of [31], the Howe-Moore theorem, and asymptotics for matrix coefficients. More precisely:

Non-tempered unitary representations may be expressed as Langlands quotients  $J(\sigma, z)$ , where  $\sigma$  is a (finite-dimensional) representation of the Levi of the  $\mathbb{R}$ -parabolic subgroup, and  $z \in \mathbb{R}$ . We normalize matters so that  $\sigma$  trivial,  $z = 1$  corresponds to the trivial representation.

Suppose we are given a – not necessary irreducible – unitary representation  $V$  of  $G$  that possesses a spectral gap. The definition of “spectral gap” easily implies that there exists  $\delta > 0$  so that any irreducible constituent of  $V$  (i.e., any irreducible representation that occurs in the support of its unitary decomposition) is one of the following types:

- (1) Tempered;
- (2)  $J(\sigma, z)$  where  $\sigma$  is not the trivial representation;
- (3)  $J(\sigma, z)$  where  $\sigma$  is the trivial representation, and  $|z| < 1 - \delta$ .

By the asymptotic expansion of matrix coefficients (see [30, Theorem 8.32]) we deduce the following: For any  $\delta > 0$ , there exists  $p$  so that any representation  $J(\sigma, z)$  with  $0 \leq z \leq 1 - \delta$  is  $\frac{1}{p}$ -tempered. It remains, then, to show that there exists a uniform  $p$  so that:

(C.1) Any unitary  $J(\sigma, z)$ , where  $\sigma$  is nontrivial, is  $1/p$ -tempered.

Fix  $\sigma$  nontrivial. Set  $z_0$  to be the supremum of those  $z \geq 0$  for which  $J(\sigma, z)$  is unitary. Proposition 45 and Theorem 6 of [31], together, imply that  $z_0 = z_c$ , the “critical abscissa” of [31] defined prior to (9.1), and moreover  $J(\sigma, z_0)$  is itself unitary (i.e., unitarizable). We shall prove in a moment that  $z_0 < 1$ . Assuming this, the definition of critical abscissa shows  $z_0$  belongs to a discrete subset of  $[0, 1)$ , independent of  $\sigma$ . The asymptotic expansion of matrix coefficients, again, shows (C.1).

It remains to show that  $J(\sigma, z)$  can only be unitary for  $z < 1$ . But that is, again, a consequence of the asymptotic expansion of matrix coefficients – [30, Theorem 8.32] – in combination with the Howe-Moore theorem that asserts that matrix coefficients decay at  $\infty$ . (One needs to know in addition that the leading exponent survives in the Langlands quotient, which is [30, Proposition 8.61]).

## REFERENCES

- [1] Y. Benoist and H. Oh. Effective Equidistribution of  $S$ -integral points on symmetric varieties. Preprint, [arxiv:0706.1621](#).
- [2] J. Bernstein and A. Reznikov. Andre Sobolev norms of automorphic functionals. *Int. Math. Res. Not.* 2002, no. 40, 2155–2174.
- [3] A. Borel and J. Tits. Eléments unipotents et sous-groupes paraboliques de groupes réductifs, I. *Invent. Math.* 12, 1971.
- [4] M. Burger. Horocycle flow on geometrically finite surfaces. *Duke Math. J.* 61 (1990), no. 3, 779–803.
- [5] M. Burger and P. Sarnak. Ramanujan duals. II. *Invent. Math.*, 106(1):1–11, 1991.
- [6] L. Clozel. Démonstration de la conjecture  $\tau$ . *Invent. Math.*, 151(2):297–328, 2003.
- [7] L. Clozel, H. Oh, and E. Ullmo. Hecke operators and equidistribution of Hecke points. *Invent. Math.* 144(2):327–351, 2001.
- [8] M. Cowling. Sur les coefficients des représentations unitaires des groupes de Lie simples. In “Analyse Harmonique sur les Groupes de Lie, II.”
- [9] M. Cowling, U. Haagerup, and R. Howe. Almost  $L^2$  matrix coefficients. *J. Reine Angew. Math.*, 387:97–110, 1988.
- [10] S. Dani. On orbits of unipotent flows on homogeneous spaces. II. *Ergodic Theory Dynam. Systems* 6 (1986), no. 2, 167–182.
- [11] S.G. Dani and G.A. Margulis. Values of quadratic forms at primitive integral points. *Invent. Math.* 98 (1989), no. 2, 405–424.
- [12] S.G. Dani and G.A. Margulis. Orbit closures of generic unipotent flows on homogeneous spaces of  $SL(3, \mathbb{R})$ . *Math. Ann.* 286 (1990), no. 1-3, 101–128.
- [13] S. Dani and G. Margulis. Asymptotic behaviour of trajectories of unipotent flows on homogeneous spaces. *Proc. Indian Acad. Sci. Math. Sci.* 101 (1991), no. 1, 1–17.
- [14] S. Dani and G. Margulis. Limit distributions of orbits of unipotent flows and values of quadratic forms. *Adv. Soviet. Math.* 16, AMS, 1993.
- [15] S. Dani and J. Smillie. Uniform distribution of horocycle orbits for Fuchsian groups. *Duke Math. J.* 51 (1984), no. 1, 185–194.
- [16] W. Duke, Z. Rudnick and P. Sarnak. Density of integer points on affine homogeneous varieties. *Duke Math. J.* 71 (1993), 143–179.
- [17] M. Einsiedler. Ratner’s theorem on  $SL(2, \mathbb{R})$ -invariant measures. *Jahresber. Deutsch. Math.-Verein.* 108 (2006), no. 3, 143–164.
- [18] M. Einsiedler, E. Lindenstrauss, P. Michel, and A. Venkatesh. Distribution of periodic torus orbits on homogeneous spaces. Preprint, [arxiv: math/0607815](#).
- [19] A. Eskin and C. McMullen. Mixing, counting, and equidistribution in Lie groups. *Duke Math. J.* 71 (1993), no. 1, 181–209.
- [20] J. Ellenberg and A. Venkatesh. Local-global principles for representations of quadratic forms. Preprint, [arxiv: math/0604232](#).
- [21] A. Eskin and H. Oh. Ergodic theoretic proof of equidistribution of Hecke points. *Ergodic Theory Dynam. Systems* 26 (2006), no. 1, 163–167.
- [22] A. Eskin and H. Oh. Representations of integers by an invariant polynomial, and unipotent flows. *Duke Math. J.* 135 (2006), no. 3, 481–506.
- [23] L. Flaminio and G. Forni. Invariant distributions and time averages for horocycle flows. *Duke Math. Journal*, 119: 465–526, 2003.
- [24] A. Gorodnik, F. Maucourant, and H. Oh. Manin’s conjecture on rational points of bounded height and adelic mixing. Preprint, [arxiv: math.NT/0601127](#)
- [25] A. Gorodnik and B. Weiss. Distribution of lattice orbits on homogeneous varieties. Preprint, to appear in GAFA, [arxiv: math/0407345](#).
- [26] D. Kleinbock and G. Margulis. Logarithm laws for flows on homogeneous spaces. *Invent. Math.*, 138(3):451–494, 1999.

- [27] D. Kleinbock and G. Margulis, Flows on homogeneous spaces and Diophantine approximation on manifolds. *Ann. of Math.* (2) 148 (1998), no. 1, 339–360.
- [28] D. Kleinbock, N. Shah, and A. Starkov. Dynamics of subgroup actions on homogeneous spaces of Lie groups and applications to number theory. *Handbook of dynamical systems* Vol. 1A, 813–930, North-Holland, Amsterdam, 2002.
- [29] A. Knapp. Lie groups beyond an introduction. Birkhäuser, 2002.
- [30] A. Knapp. Representation theory of semisimple groups. Princeton University Press, 2001.
- [31] A. Knapp and E. Stein. Intertwining operators for semisimple groups. *Annals of Math.* 93 (1971).
- [32] E. Leuzinger. Kazhdan’s property (T),  $L^2$ -spectrum and isoperimetric inequalities for locally symmetric spaces. *Comment. Math. Helv.*, 78 (2003), no. 1, 116–133.
- [33] J. Li, D. Jiang and S. Zhang. Periods and distribution of cycles on Hilbert modular varieties. *Pure Appl. Math. Q.* 2 (2006), no. 1, 219–277.
- [34] D. Luna. Sur certaines opérations différentiables des groupes de Lie. *Amer. J. Math.*, 97:172–181, 1975.
- [35] B. Malgrange. Ideals of differentiable functions. Oxford University Press, 1966.
- [36] G.A. Margulis. The action of unipotent groups in a lattice space. *Mat. Sb. (N.S.)* 86(128) (1971), 552–556.
- [37] G.A. Margulis. Discrete subgroups of semisimple Lie groups. *Ergebnisse der Mathematik und ihrer Grenzgebiete*, Springer-Verlag, 1989.
- [38] G.A. Margulis. Formes quadratiques indéfinies et flots unipotents sur les espaces homogènes. *C. R. Acad. Sci. Paris Sr. I Math.* 304 (1987), no. 10, 249–253.
- [39] G.A. Margulis. Discrete subgroups and ergodic theory. *Number theory, trace formulas and discrete groups*, (Oslo, 1987), 377–398, Academic Press, Boston, MA, 1989.
- [40] G.A. Margulis. Indefinite quadratic forms and unipotent flows on homogeneous spaces. *Dynamical systems and ergodic theory* (Warsaw, 1986), 399–409, Banach Center Publ., 23, PWN, Warsaw, 1989.
- [41] G.A. Margulis. On some aspects of the theory of Anosov systems. *Springer monographs in mathematics*, 2004.
- [42] G.A. Margulis and G. Tomanov. Invariant measures for actions of unipotent groups over local fields on homogeneous spaces. *Invent. Math.* 116 (1994), no. 1-3, 347–392.
- [43] G. A. Margulis and G. Tomanov. Measure rigidity for almost linear groups and applications. *J. Anal. Math* 69 (1996), 25–54.
- [44] P. Michel and A. Venkatesh. Equidistribution, L-functions and ergodic theory: on some problems of Yu. V. Linnik. *International Congress of Mathematicians. II* (2006) 421–457. Eur. Math. Soc., Zürich
- [45] S. Mozes, N. Shah. On the space of ergodic invariant measures of unipotent flows. *Ergodic Theory Dynam. Systems* 15 (1995), no. 1, 149–159.
- [46] A. Nevo. Spectral transfer and pointwise ergodic theorems for semi-simple Kazhdan groups *Math Research Letters*, 1998.
- [47] H. Oh. Uniform pointwise bounds for matrix coefficients of unitary representations and applications to Kazhdan constants. *Duke Math. J.* 113 (2002), no. 1, 133–192.
- [48] J.-F. Quint, Propriété de Kazhdan et sous-groupes discrets de covolume infini. *Trav. Math.*, XIV, 143–151, Univ. Luxemb., Luxembourg, 2003.
- [49] M. Ratner. Rigidity of horocycle flows. *Ann. of Math.* 115 (1982), 597–614.
- [50] M. Ratner. Factors of horocycle flows. *Ergodic Theory Dynam. Systems* 2 (1982), no. 3-4, 465–489.
- [51] M. Ratner. Horocycle flows, joinings and rigidity of products. *Ann. of Math.* (2), 118(2):277–313, 1983.
- [52] M. Ratner. Rigidity of time changes for horocycle flows. *Acta. Math.* 156, 132, 1986.
- [53] M. Ratner. On measure rigidity of unipotent subgroups of semisimple groups. *Acta Math.* 165 (1990). 229–309.

- [54] M. Ratner. Strict measure rigidity for unipotent subgroups of solvable groups. *Invent. Math.* 101 (1990), 449–482.
- [55] M. Ratner. Raghunathan’s topological conjecture and distributions of unipotent flows. *Duke Math. J.* 63 (1991), no. 1, 235–280.
- [56] M. Ratner. On Raghunathan’s measure conjecture. *Ann. of Math. (2)*, 134(3):545–607, 1991.
- [57] M. Ratner. Raghunathan’s conjectures for  $SL(2, \mathbb{R})$ . *Israel. J. Math.* 80, 131, 1992.
- [58] M. Ratner. Raghunathan’s conjectures for Cartesian products of real and  $p$ -adic Lie groups. *Duke Math. J.* 77 (1995), no. 2, 275–382.
- [59] R. W. Richardson. Conjugacy classes in Lie algebras and algebraic groups. *Annals of Mathematics*, 86 : 1–15, 1967.
- [60] A. Strömbergsson. On the uniform equidistribution of long closed horocycles. *Duke. Math. Journal* 123: 507–547, 2004.
- [61] A. Venkatesh. Sparse equidistribution problems, period bounds, and subconvexity. preprint, [arxiv:math/0506224](https://arxiv.org/abs/math/0506224).
- [62] N. Wallach. Real reductive groups I. vol 132 of *Pure and Applied Mathematics*, Academic Press, 1988.
- [63] B. Weisfeiler. On one class of unipotent subgroups of semisimple algebraic groups. [arxiv: math/0005149](https://arxiv.org/abs/math/0005149).
- [64] D. Witte. Rigidity of some translations on homogeneous spaces. *Bull. Amer. Math. Soc. (N.S.)* 12 (1985), no. 1, 117–119.
- [65] D. Witte. Rigidity of some translations on homogeneous spaces. *Invent. Math.* 81 (1985), 1–27.
- [66] D. Witte Morris. Ratner’s theorems on unipotent flows. *Chicago Lectures in Mathematics*. University of Chicago Press, Chicago, IL, 2005. xii+203 pp.
- [67] Witte, Dave, Yukie, Akihiko and Zierau, Roger. Prehomogeneous vector spaces and ergodic theory. II. *Trans. Amer. Math. Soc.* 352 (2000), no. 4, 1687–1708.
- [68] Yukie, Akihiko. Prehomogeneous vector spaces and ergodic theory. I. *Duke Math. J.* 90 (1997), no. 1, 123–147.